

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 19, 2000, 23:39:15 ; Search time 4256.96 Seconds
(without alignments)
1508.199 Million cell updates/sec

Title: US-09-339-159-1
Perfect score: 1470
Sequence: 1 ttgaataatggttttaaaaa.....ttgataatgtgattgtagaag 1470

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_v1:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pr4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

BEST AVAILABLE COPY

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pr5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_v11:*
73: gb_v12:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pr6:*
92: gb_pr7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	460.8	31.3	1551	1	AB007123	AB007123 Escherichia coli
2	242	16.5	1530	2	D86329	D86329 Vibrio cholerae
3	134	9.1	4567	1	AF163837	AF163837 Caldococcus
4	131.8	9.0	2155	2	CDCMANABD	M56063 C.saccharifaciens
5	131.8	9.0	4977	2	CDCMANA	L01257 Caldococcus
6	125	8.5	5284	2	CSU16308	U16308 Caldococcus
7	109.4	7.4	1461	5	E02075	E02075 Genomic DNA
8	109.4	7.4	1939	2	BACMANN	M31797 Bacillus sp
9	94.4	6.4	25970	2	SC2H4	AL031514 Streptomyces
10	89.2	6.1	1881	74	STMBMANASE	M92297 Streptomyces
11	81.6	5.6	837	74	TFU6227	AJ006227 Thermotoga
12	52.4	3.6	14922	2	SCF73	AL121746 Streptococcus

C	13	48.6	3.3	7218	5	166494	I66494 Sequence 14
C	14	45.8	3.1	111071	53	AC024591	AC024591 Homo sapi
C	15	45.2	3.1	35049	33	CEIK03E6	U53375 Ceenorhabdi
C	16	45	3.1	175300	40	AC012459	AC012459 Homo sapi
C	17	44.6	3.0	4807	7	AB006362	AB006362 Candida a
C	18	44.6	3.0	143092	56	AC051641	AC051641 Homo sapi
C	19	44	3.0	14235	73	H267264	U67264 Helicoverpa
C	20	44	3.0	147123	54	AC027030	AC027030 Homo sapi
C	21	43.2	2.9	2138	8	CA015801	U15801 Candida alb
C	22	42.8	2.9	14867	8	HS10189	AL035455 Human DNA
C	23	42.8	2.9	251206	68	AL354834	AL354834 Homo sapi
C	24	42.4	2.9	1063	72	HA067261	U67261 Helicoverpa
C	25	41.8	2.8	42539	66	AC078830	AC078830 Staphyloc
C	26	41.8	2.8	47963	60	AC074320	AC074320 Staphyloc
C	27	41.8	2.8	54249	66	AC078934	AC078934 Staphyloc
C	28	41.8	2.8	55418	53	AC023592	AC023592 Staphyloc
C	29	41.8	2.8	214412	43	AC018695	AC018695 Homo sapi
C	30	41.6	2.8	17660	48	SC9723	Z48951 S.cerevisia
C	31	41.6	2.8	38779	48	SCU31900	U31900 Saccharomyc
C	32	41.6	2.8	107642	12	AF125313	AF125313 Mus muscu
C	33	41.6	2.8	165536	48	SCCHXVI	Z71255 S.cerevisia
C	34	41.4	2.8	3734	8	ATRNABELI	X97970 A.thaliana
C	35	41.4	2.8	81493	8	ATB1KBGEN	X98130 A.thaliana
C	36	41.4	2.8	82646	7	AB028611	AB028611 Arabidops
C	37	41.4	2.8	182593	39	AC011209	AC011209 Homo sapi
C	38	40.8	2.8	1302	11	AF090946	AF090946 Homo sapi
C	39	40.8	2.8	56804	49	H577N19	Z98886 Human DNA S
C	40	40.8	2.8	155740	53	AC025003	AC025003 Homo sapi
C	41	40.8	2.8	191126	41	AC013444	AC013444 Homo sapi
C	42	40.6	2.8	3804	48	YSCMTG04	L38888 Saccharomyc
C	43	40.6	2.8	3804	48	YSCMTG04	J01485 Yeast (S.ce
C	44	40.2	2.7	2103	57	TW049024	U49024 Tegetiula
C	45	40.2	2.7	44101	60	AC073960	AC073960 Homo sapi

ALIGNMENTS

RESULT	1	AB007123	1551 bp	DNA	BCT	28-APR-1998
LOCUS	AB007123	Bacillus circulans gene for mannanase, complete cds.				
DEFINITION	AB007123	Bacillus circulans gene for mannanase, complete cds.				
ACCESSION	AB007123.1	GI:3090433				
VERSION	AB007123.1	GI:3090433				
KEYWORDS	guar gum-degrading enzyme; mannanase.					
SOURCE	Bacillus circulans DNA.					
ORGANISM	Bacillus circulans					
REFERENCE	1	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.				
AUTHORS	Yoshida, S.					
TITLE	Direct Submission					
JOURNAL	Submitted (08-SEP-1997) to the DDBJ/EMBL/GenBank databases. Seq11					
REFERENCE	Yoshida, S., Sako, Y. and Uchida, A.					
AUTHORS	Cloning, sequence analysis, and expression in Escherichia coli of a					
TITLE	gene coding for an enzyme from Bacillus circulans K-1 that degrades					
JOURNAL	guar gum					
MEDLINE	Biosci. Biotechnol. Biochem. 62 (3), 514-520 (1998)					
FEATURES	98233274					
source	Location/Qualifiers					
CDS	1..1551					
	/organism="Bacillus circulans"					
	/db_xref="taxon:1397"					
	1..1551					
	/function="guar gum-degrading enzyme"					
	/codon_start=1					
	/transl_table=1					
	/product="mannanase"					
	/protein_id="BA025878.1"					

QY	85	acagctaatgcaaatccggaatttatgtaagcgggtaccactctatacgatgcaatgga	144	/db_xref="GI:3090434"
DB	91	AAAGCAGATGCTGCAAGCGGATTTATGTAGCGGTACCAATATTGATGATCAAGGA	150	/translation="MGWFLVLRKMLIAFAVFLMFMSQTQLNKAHAGFFWGGTX
QY	145	aaccattgtaagtagagggattacacatgsgaagcattgataagaccaggaact	204	LIDATGQFVWVGHVHAHWTWKDLSAIPAIKATGANTRIYLANHKTLDVNTV
DB	151	CAACATTTGATGCGAGAGTCAATATCGCACATGATGTAAGATCAACTATCC	210	NNILICEONKLIIVLEVHDATGDSLSDANVYGIKSLIGEDVITIANE
QY	205	actgcaattgagggattgcaataacggtgtaatacgttcggattgltatcgtat	264	MYGTGWDVAANVANGKQAKPLRNKGLHTLIVBAGGGOVDSKXNCTEVLNADPLK
DB	211	ACCGCAATACGACGATGCTGTAACAGTGCACACAGATGATGATGATGCGGAA	270	MYTFSIMVEYAGNASTKSTVNDVGNKLIATGCGGGOVDSKXNCTEVLNADPLK
QY	255	ggggagcaatgacaaagatgacatcgcagtaagaacccattatctttgacgaa	324	KGVGWLMSKMGNSDIAIDMTDMAGNSLITSGNTVAGSNIKATSVLSGIFGV
DB	271	GGACCAATATGACGCTTGATGATGATTAACACCGTCAACAAATATTCACCCCTGTGA	330	TPISSTPTSTPTSTPT
QY	325	gatacatctggttctcttctgaagtcattgatgtaacggttatgattcatcgt	384	EMKATGATLKADVLSQNSNTHSLYITSONLSGKSLKATVHAMNGNNGIYATL
DB	331	CAAAACAACATTAATTCGGTTTGGAGTACATACCGTCAAGACGATAGCTTCC	390	YVKTSGMTWDSGSENLIOISNDGIITLISGISLSSVKEIGVEFASNSGOSAI
QY	385	tcgctcaatcgctgctgttattatgtaattgtaagaatgtaagtcattatggaaga	444	YVDSVSLIO"
DB	391	GATTTAGCAACGCGCTTAATTAATGATGATGATGATGATGATGATGATGATGATG	450	
QY	445	gataccgcatattatattatggaatggaatggtttgctggtggaaggagatcgtg	504	
DB	451	GACGCTGATCATTAATTAATGATGATGATGATGATGATGATGATGATGATGATG	510	
QY	505	gctgacgggtataaacaagaacatcccggaattggaacggttctaaacatcacttg	564	
DB	511	GCTAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	570	
QY	565	atggtatgctgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	624	
DB	571	ATTGTTAGCTCTCGGTGATGATGATGATGATGATGATGATGATGATGATGATG	630	
QY	625	gttttaattgctgacccctcaacgaatatacaatggttttcgattcatatgataatga	684	
DB	631	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	690	
QY	685	gtgtgtaattgctgcaatgctgctgctgctgctgctgctgctgctgctgctgctg	744	
DB	691	GGGGGCAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	750	
QY	745	ttaattgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	804	
DB	751	CTGATTAATGCGGATTTGATGATGATGATGATGATGATGATGATGATGATGATG	810	
QY	805	atgagctatctgaacaagaaggatggtggtggtggtggtggtggtggtggtggtg	864	
DB	811	ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	870	
QY	865	ccagaatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	924	

Query Match 31.3%; Score 460.8; DB 1; Length 1551;
 Best Local Similarity 58.9%; Pred. No. 1.9e-97;
 Matches 859; Conservative 0; Mismatches 527; Indels 72; Gaps 1;

```

Db      871 AGTATTTGGCTTATCTCGATATGACAAATGATGGCGTGTACCTCCCTACCTCGTTC 930
Qy      925 ggaataacataatgtaatggtccataatggttaagaagaactcgagattgaacacggtt 984
Db      931 GGTATACCGTAGTAAATGCGAGTAAACGCGCATTAAGCACTTCTGTATACCGCATT 990
Qy      985 tttaacaggctgagagatctgtagtgaagactctcgcaactc----- 1027
Db      991 TTGGAGGTTGTTACGCCAACCTCAAGCCCTACTCTACCTACCTACCTACGCCAACCTCA 1050
Qy      1028 -----tttat 1032
Db      1051 ACTCTACTCTCAGCGCAAGTCCGACCCGAGTCCAGGTAATACGGGACGATCTTATAT 1110
Qy      1033 gattttgaagtagtataatgcaagatgagactggaagtagcttgaacgaggtcttgggtc 1092
Db      1111 GATTTCGAAAGAGCAAGCAAGGCTGTGCGGAAACAATATTTCCGGAGGCCCATGGGTC 1170
Qy      1093 gtagacagagtggtctctcaagaagagatctctttaaagaagagatcaattgctcgtca 1152
Db      1171 ACCAATGATGGAAGAAAGCAAGCGGACGCAAACTCTCAAGCCGATGCTCTTACCAATCC 1230
Qy      1153 aattcacacattactacatgataatcaaatgcaatgctcttcaacagcagaatagtagata 1212
Db      1231 AATTCACAGCAGTACTATATATATATACCTCTAATAATCTGTCTGAGAAAAAGAGTCTG 1290
Qy      1213 caagctactgttaaacatgcaaatgagggaagtgctgtaatggaatgactgagcgtctt 1272
Db      1291 AAGAGCAAGGTTTAAAGCAATGCGAACGAGGCGCAATATCGGCACAGGGATTTATGCAAACTA 1350
Qy      1273 tatgtgaagaagagacatggtatatactgtaactgtaagagcttgctgagatgaagct 1332
Db      1331 TACGTAAGACCGGCTCCGGGTGACATGTAAGTCCGGAGAGAAATCTGTATCAATCA 1410
Qy      1333 tcaatctggaacacagcctatctctagattatcaaatgccaatcttctcaagaag 1392
Db      1411 AACGACGATACCATTTTACACTATCCCTCAGCGCATTTGCAATTTGCCCATCACTCAA 1470
Qy      1333 gaaatgtgagtcagttcccaatcagcagtgtagtagtaggaacaaatcgatataatt 1452
Db      1471 GAAATTTGGGTAGAAATTCGGCCCTCTCAACAGTACGAGGCAATCTATTTATGTA 1530
Qy      1453 gataatgtgattgtaga 1470
Db      1531 GATAGTGTACTCTGCA 1548

RESULT 2
D86329 1530 bp DNA BCT 19-MAR-1998
DEFINITION Vibrio sp. gene for beta-1,4-mannanase, complete cds.
ACCESSION D86329.1 GI:2978250
VERSION beta-1,4-mannanase; manA.
KEYWORDS Vibrio sp. (strain:MA-138) DNA.
SOURCE Tamaru, Y.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 1530)
AUTHORS Tamaru, Y.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1996) to the DDBJ/EMBL/GenBank databases. Yutaka
Tamaru, Mie University, Faculty of Biosources, Utilization of
Marine Products, 1515 Kamihama, Tsu, Mie 514, Japan
(E-mail: s9tamayemarie.bio.mie-u.ac.jp, Tel:0592-31-9561,
Fax:0592-31-9557)
2 (sites)
REFERENCE Tamaru, Y., Araki, T., Morishita, T., Kimura, T., Sakia, K. and
AUTHORS Ohnishi, K.
TITLE Cloning, DNA sequencing, and expression of the beta-1,4-mannanase
JOURNAL gene from a marine bacterium, Vibrio sp. strain MA-138
FEATURES
location/Qualifiers
1..1530

```

```

/organism="Vibrio sp."
/strain="MA-138"
/db_xref="taxon:678"
-35_signal
-10_signal
RBS
gene
CDS
136..141
149..1336
/gene="manA"
149..1336
/gene="manA"
/codon_start=1
/transl_table=1
/product="beta-1,4-mannanase"
/protein_id="P42518.1"
/db_xref="GI:2978251"
/translation="MKFTKAIISLLEITMASCAGFYVSNGLYEANGSAFKIRGIN
HAHWYDKLSVALSGIAAGAVNRVVLNSGRMTKNDVSDVINILKANNLILAI
LEVHDTTGKESSEASLDSAADWIELKELIGOEYVILINLEPENNDAVAT
NDHSAIORLSAGINHTIMVDAPNMGDMKFWLNAOFVFNEDPKLITFSTVST
VYSXNVDNIYISFTNNGLYLVIGERASTHKGADNDGSIKESLSTCYTFRS
GNDTISDLIDIVNMDNNSTSTMGVNLINGONGIKSTLATVTCNDKNDPSCTV
PICSSAVDPDGDGWMENNOCIVODSDTAPNGPYPCQESSDPDGDGWMENNNS
MCC"
BASE COUNT 469 a 211 c 333 g 517 t
ORIGIN
Query Match 16.5% Score 242; DB 2; Length 1530;
Best Local Similarity 57.2% Pred. No. 2.4e-46;
Matches 525; Conservative 0; Mismatches 375; Indels 18; Gaps
Qy 93 tgcaatccgattttatgtaagcggtaacactctatagatgcaatggaacccatt 152
Db 202 TGCTCATCGGGTTTATGATCAAAATGTCCTTATATGAGCAAGATGTAGTGCCTT 261
Qy 153 tgaatgagaggaatcaaccatgagcagcatgataaagaccagcaactactgcaac 212
Db 262 TAAATTTAGAGGATTAATTCATGCATCTTGTGATACAGCAAGATTAAACGTGGCTT 321
Qy 213 tgaaggatctcaaatcagcgtgtaatacagtcggtccggaatggtatctgtagtgggaca 272
Db 322 AAGTGTATGCTGCCACAGAGCAATACGTGACGGGTGTGACTAAGTAAATGATTATCG 381
Qy 273 atgacaaagaatgacacacacagtaagaaccttacccttaaggagaataatca 332
Db 382 GTGCAAAATAATGACGTTAGATGATGATCACTAATTAATTAATTAAGCAAAACAA 441
Qy 333 ttgtgtcgtctctgaatgacatgacacggtatg-----atccattgc 383
Db 442 CCTTATACCTATATTAAGAAATTCATGACACTACTGATGAGAGAAAGATAGTGTGC 501
Qy 384 ttgcctcaatcgtgctgtgattatgattgtaataatgaagagtgcttcaattggaagga 443
Db 502 TAGTTAGATTCTCGCGGTACTGATGATGATGATGATGATGATGATGATGATGATGAT 561
Qy 444 agatacgcattatataatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatg 500
Db 562 AGATATGTCATCATTAACCTACGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 621
Qy 501 ttggctcagcggatataaacaacaccccgcatgtagtaacgcggtcttaaacatcac 560
Db 622 TTGGTTAATGACCATGTCAGTGCATATCAACGATTAAGAAAGTGCAGGATTAACACAC 681
Qy 561 ctgtagttagatgctgctgggggtggaacaa---ttcccaatcgaatcgaatgtag 617
Db 682 GATTATGCTGATGCTCAAACTGGGCGGACAGACTGGAAGGCGTTATTCGTAATATATGC 741
Qy 618 aagaagaatttttaattgctgacccctcaacgaataacaaatgcttccgattcataatgata 677
Db 742 TCAATTGCTTTTAAATTTGACCCCTTAATTAATTAATTAATTAATTAATTAATTAAT 801
Qy 678 atagcagtgtagatgcatgcaaggtcgaataattgacagaggtcgaataatgata 737

```

Db 802 GGTATAGTACTTACAAATC---AGTCAGACTACATCTCATATTACAAACAATGG 858
 QY 738 cctcgcatatgcatctgtgtaatttgagcaccgctacatacaaatggtgacgtcgatgaagc 797
 Db 859 CTATAGTGGTATTCGGGAGGACTTGCCTTCAACATCAATAAGGGCGGAGATGTGATGAAGG 918
 QY 738 aacgaattatagcatcttctgaacaaagagagtggtggtgtgcgtggtcattggaag 857
 Db 919 TCAATTAATAGGAACGTTGAGAAACGTTAAGTTAGTTATATCGGTGGTCTGCTGACAG 978
 QY 858 gaagggccagagatggagatattagaccttcgagatgattggcgctggaataaccttac 917
 Db 979 AATGATGTTACACATCCGATTTGGATTTGTTAATTTGGGACAAATAATTCGATAG 1038
 QY 918 agcttgggaaatacaatagtagtgcatactggttaagaagaacctcgagaattag 977
 Db 1039 CACTTGGGCTAATGTTTAATTAATGCTCAAAATGATTAATCAATCAGCTGACG 1098
 QY 978 caacgttttacaagtgg 995
 Db 1099 TACGGTGTACTGTGG 1116

RESULT 3
 AF163837 4567 bp DNA BCT 08-FEB-2000
 LOCUS Caldicellulosus multidiomannanase
 DEFINITION precursor (mana) gene, complete cds; and unknown genes.
 ACCESSION AF163837
 VERSION AF163837.1 GI:6651325
 KEYWORDS Caldicellulosus cellulovorans.
 SOURCE Caldicellulosus cellulovorans.
 ORGANISM Caldicellulosus cellulovorans;
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Caldicellulosus.
 REFERENCE 1 (bases 1 to 4567)
 AUTHORS Sunna, A., Gibbs, M.D., Chln, C.W.J., Nelson, P.J. and Bergquist, P.L.
 TITLE A gene encoding a novel multidomain beta-1,4-mannanase from
 Caldicellulosus cellulovorans and action of the recombinant enzyme on
 kraft pulp
 JOURNAL Appl. Environ. Microbiol. 66 (2), 664-670 (2000)
 MEDLINE 20120520
 REFERENCE 2 (bases 1 to 4567)
 AUTHORS Sunna, A., Gibbs, M.D. and Bergquist, P.L.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUN-1999) Biological Sciences, Macquarie University,
 North Ryde, New South Wales 2109, Australia
 FEATURES
 source
 1..4567
 /organism="Caldicellulosus cellulovorans"
 /db_xref="taxon:74586"
 <1..513
 /note="orf1"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF22273.1"
 /db_xref="GI:6651326"
 /translation="PTPTPTPTSAAPTPTPSAGSLVVOYRAADTNAGDNOLKPHFRI
 VNRGTSVPLSELTIRYVTVGDKPQVNCMAQVGCNSVNGSEFVKLSTGTGADY
 IETFTSGAGSLAAGASGSDIOYRINKNDMTVNEANDYSYPTKTSFDMRWVLYR
 NGOLIMGVEP"
 misc-feature
 1..60
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 misc-feature
 61..510
 /note="Region: cellulose-binding domain type IIb"
 /evidence="not_experimental"
 620..3412
 /gene="mana"
 /gene="mana"
 620..718
 /gene="mana"
 /evidence="not_experimental"
 sig_peptide
 /evidence="not_experimental"

CDS
 620..3412
 /gene="mana"
 /codon_start=1
 /transl_table=11
 /product="multidiomannanase precursor"
 /protein_id="AAF22273.1"
 /db_xref="GI:6651327"
 /translation="MNRRLIARLSGLMAVLIAMLAAYVYKPEPAHAGMVPATRT
 YACVVGKHNKNGDLMNINPACLDALATSGVNPQPNMGLNLSNAGGRREIIPGK
 LCGPTASFDGMNARDIMWTRILQPGCATTVRMANAPHPGTIVLVTDGHPQPL
 KMDLEPTNSQVNTNPINSSGDAEYSQVLPKNGHIIYIMLWNSDSEATYN
 CSVYFGSGPIAEFGDPREGGTMIPPEGGTYTPPTPTPTPTPTPTPTPTPTPT
 TPT
 TPT
 STAOEYFCMAOIGCSNIRAPVSLSPVSGADSYIELFTGSLIPAGNTEIONR
 IHTNNMKNVETPDMSVNGAQTGWGPSTRITLYRNGVLVWTEPGGSSPPPTPT
 PT
 GAAVAVTLVSNRGKWKIPSEVADITISQARTLGRVAVLEVDHTTGSGDAACSM
 TAVNWIETLNKVLAGEVEVIVNIGNEPQNNYQMWVDTRNAVALRAGLINTIM
 VDAPNMGDWSFTMRDNPATIEFNADPQRLNLFSLHMYGVDTAEVQSYTESVNGI
 PLVIGEGHMSDGPNEOAIYOAKOYNGLFSGMSGSGVEYLDVNTNANSP
 TANGTPT
 KPHRIYNRGTSVPLSELTIRYVTVGDKPQVNCMAQVGCNSVNGSEFVKLSTGR
 TGADYIETFTSGAGSLAAGASGSDIOYRINKNDMTVNEANDYSYPTKTSFDMRW
 VLYRNGOLVWGEV"
 719..3409
 /gene="mana"
 /product="multidiomannanase"
 1349..1480
 /gene="mana"
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 1481..1927
 /gene="mana"
 /note="Region: cellulose-binding domain type IIb"
 /evidence="not_experimental"
 1928..2032
 /gene="mana"
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 2033..2884
 /gene="mana"
 /note="Region: beta-1,4-mannanase catalytic domain"
 /evidence="experimental"
 2885..2959
 /gene="mana"
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 2960..3409
 /gene="mana"
 /note="Region: cellulose-binding domain type IIb"
 /evidence="not_experimental"
 3894..4400
 /note="orf3"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF22275.1"
 /db_xref="GI:6651328"
 /translation="MGASGVSPRTAONGYGLKARCOKRCVGRNLGRLFVVDDEF
 RFFGARRLDREFSRNORLCEANFESGYTTPQPPDKRIIRENGLFLVROAST
 GPKRIELIADROARARQASDQPOLKNAVILHFDIMHPYGERMDPEFHFIEKQNS
 RSLIRYV"
 BASE COUNT 1040 a 1276 c 1439 g 812 t
 ORIGIN
 Query Match 9.1%; Score 134; DB 1; Length 4567;
 Best Local Similarity 51.9%; Pred. No. 4e-21;
 Matches 408; Conservative 0; Mismatches 360; Indels 18; Gaps 4;
 QY 158 tgaagagattaacatggtgacgtatgaagaccagcaactctgcaattgaag 217
 Db 2070 TGGTTGATTTATACACCGCATGCGGTATCGGACGCGCTGATGCTTGCAGG 2129

Db	1298	ATTTGTGAATTATAAACATTGGTAATGACCGCTAATGGGCAACAATACATCAAAACTGGG	1357
OY	506	ctcagcggtataaacaagcaatcccgcgatttcgctcaagccggtctaaacatacccttga	565
Db	1358	TTAATGACACGAAAGAACGCTATTAAAGACACTTAGAGATGCAGAGATTCAAGCACAGATAA	1417
OY	566	tggtagatgcctgcggggttgggagc--aaattcccaactcgaattcatgattatggaag	622
Db	1418	TGGTGGATGCGCCGAACCTGGGGTCAGAGATTGGCTAATACATAGAGATATGCGCCAGA	1477
OY	623	aagttttaaactgtgacctcaacgnaatacaaatgttttcgattcatatgtaataatg	682
Db	1478	GCATTAATGGAGACAGATCCCTCGCAATTTGGTATTTCGATTCAATAGTAATG--GCG	1534
OY	683	caagttgtaatacgcgcgaagttcgcactaataattaccagagtttcttaacaagacctg	742
Db	1535	TATACATATACAGACAGCAAGAGTGGAAAGATACATAAATCATTTGGTGTATAGGGGTATC	1594
OY	743	cattagtcattgttgaattttggagacacgcgtcatacaaatgtgtacgttcgattgaagaaga	802
Db	1595	CATTGGTATTATGGAGATTGGATTCGACATCAGACACAAATGGTGAACCTGTATGGAAGAGCTA	1654
OY	803	ttatgagctatctcgaacaagaagaggttggtgtgttcgcgttgatctgaaaggaaacg	862
Db	1655	TTTGTCAGGTATGCAAACACATACAAAGTATAGATTATTATTGTTGGTGTGCTGTGGAAATT	1714
OY	863	gcccagaatvggagtaatttgaccttcgaatgatttgggctcgtgaaataaacttaacgctt	922
Db	1715	CGAGCTATTGTTGGGTATTATTGGACATGTGTAAACAACTGGACCCCAATTAATCCAACTCCAT	1774
OY	923	gggggaaa 929	
Db	1775	GGGGACA 1781	

RESULT	5			BCT	26-APR-1993
CDCMANA					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
FEATURES					
source					
gene					
CDS					

```

CDCMANA      4977 bp    DNA
Caldocellum saccharolyticum beta-mannanase/endoglucanase (mana)
gene, complete cds.
L01257
L01257.1 GI:144290
beta-mannanase; endo-1,4 beta-mannanase; endoglucanase; mana gene
Caldocellum saccharolyticum (library: lambda NBP2) DNA.
Caldicellulositractor saccharolyticus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Thermocaneorobacter group; Caldicellulositractor.
1 (bases 1 to 4977)
Glbbs,M.D., Saul,D.J., Luthi,E. and Bergquist,P.L.
The beta-mannanase from Caldocellum saccharolyticum is part of a
multidomain enzyme
J Appl. Environ. Microbiol. 58, 3864-3867 (1992)
93119139

Location/Organism
1..4977 /organism="Caldicellulositractor saccharolyticus"
/db_xref="taxon:44001"
841..4836 /gene="mana"
841..4836 /gene="mana"
/gene="mana"
/gene="mana"
/EC_number="3.2.1.78"
/note="endoglucanase EC_Number= 3.2.1.4"
/codon_start=1
/transl_table=1
/product="beta-mannanase"
/protein_id="AAI1887.1"
/db_xref="GI:144291"
/transl_table="MRLTKIRKKMLSVLCVTFELNLIIFIANVTILPKVGASNSN
VKVDSTLLIGTNHAGWYRDLDALRGIRSGKMSVRYVLSNGYRWTKIPASEVV
ITISRSGLIKRAIIIEVHDFTGGEGAGCSLAQAVEYRKETKSVLNDGEFVIIN
NEPSGRNNYONWVPDPTNAIKALPDAGAFHTIMVDA.PNMGDQSMNTMDAQSIWE
PLRNLVSEIHMTGYNTAKSVEEYIKSFYDKGLPIYIGFGHQHIDGPDEALYR

```

BASE COUNT 1698 a 815 c 1265 g 1199 t
ORIGIN

Query Match	9.0%;	Score 131.8;	DB 2;	Length 4977;
Best Local Similarity	51.7%;	Pred. No. 1.3e-20;		
Matches 407; Conservative	0;	Mismatches 362;	Indels 18;	Gaps 4;

QY	158	tgagggagattaccatggygcacgcatgtgttaaaagaccagggcaactcgtgaattgaag	217
Db	998	TTATGGAAACCAATACGCACATTCCTGCTACAGAGATAGACTTGATACGCATTGCCGG	1057
QY	218	ggattgcgaatcccggtgctaaacggtccgattgtgylatcgtatggygggacaatgga	277
Db	1058	GAATTAGGTCATGGGGTATGAACCTGTGTAGGGGTAGTGTGTGATATGGTTTCCGTGA	1117
QY	278	caaaagatgacatccatacagtaagaacctatctctttagcggagaatacatcttgg	337
Db	1118	CGAAGATACCAAGCAAGTGAAGTGGCAAAATTATATCATTTGTCAGAACTCTTGTTTCA	1177
QY	338	ttgcgttctt---cttgaagltcatglatgctaccggtatgatattccattg-----ctt	385
Db	1178	AAGCTATTATTATTAGAAAGCGACGACACAAACAGATATGAGAAAGATGGGCGACATGTT	1237
QY	386	cgctcaatcgtgctgtgtgattgtgaatgaagaagtcttcaatttggaaaagaag	445
Db	1238	CATTGGCACAAAGCAGTGGATTTGTAAGAGAGATTAAGACGCTATTAGACGGTAAGAAAG	1297
QY	446	ataccgltcatttaatatitgcgaatgaaatgtgtgttcctgtyggagaaggatgcttgg	505
Db	1298	ATTTTTATTATTATTAACATTGGTTAAAGACCCGTATGGAAACATAATCTATCAAAACTGGG	1357
QY	506	ctgacggyratbaaacagaagaatcccgcgattggtaaagccggtctcaaacatcacttga	565
Db	1358	TTTATACACGAAGAACCTTATTAAGCACTTAGAATATCGAGATTCCAAGCACAGATTA	1417
QY	566	tgttagatgctgcgggtygggac---aattccacaatcgatcatgattatggaagag	622
Db	1418	TGCTGATGATCGCCGAACCTGGGTCAGAGATTGCTATATACATAGAGAGAAATGCCAGA	1477
QY	623	aagttttaatgtgcacccccaacgaataacaaatgttttcogattcatgatatgaaatg	682
Db	1478	GCATTAATGGAAGCAATCCGCTGCCGAATTTGGTATTTTTCGATTCAATATGTATG---GCG	1534
QY	683	caggtygtaatgcatcgcgaagtcgtaactaaatatgacggagtcttcaatcaagactcg	742
Db	1535	TATACATATACGGAAGCAAGGTGGAAGAGTACTCAATATCTATTGTGATAAAGGGTTAC	1594
QY	743	catgtgcatitgtgaatttgcacaccgctacatacaaatgttgcacgtcgaatgaaagcaaga	802
Db	1595	CATTGGTATTGGAAATTTGGACATACAGCACACAGATGGTGAACCTGATGAAGAAGCTTA	1654
QY	803	ttaatgactatcttcaacaagaagagtttgggtgtgttgcgcgtgtgcatggaagggaagc	862

Db 1655 TTGTGAGGTATGCAAAAGATGACATGATGATTTAGTGTGCTGCTGTGGAATT 1714
QY 863 gccacgaatggagattagacccttcgaatgattggctggaataaccctacagctt 922
Db 1715 CGAGCTATGTTGGATTGTGACATGATTAACACTGGAGCCCAATATCAACTCAT 1774
QY 923 ggggaaa 929
Db 1775 GGGGACA 1781

RESULT 6
LOCUS CSU16308 5284 bp DNA BCT 10-FEB-1996
DEFINITION Caldocellum saccharolyticum endoglucanase/mannanase (celC)
ACCESSION U16308
VERSION U16308
KEYWORDS pseudogene, complete cds.
SOURCE U16308.1 GI:577827
ORGANISM Caldocellum saccharolyticum.
Caldocellulosiraptor saccharolyticus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Thermoanaerobacter group; Caldocellulosiraptor.
1 (bases 279 to 4374)
Morris,D.D., Reeves,R.A., Gibbs,M.D., Saul,D.J. and Bergquist,P.L.
Correction of the beta-mannanase domain of the celC pseudogene from
Caldocellulosiraptor saccharolyticus and activity of the gene
product on Kraft pulp
Appl. Environ. Microbiol. 61 (6), 2262-2269 (1995)
95314280
REFERENCE 2 (bases 1 to 5284)
Morris,D.D.
Direct Submission
Submitted (25-OCT-1994) Daniel D. Morris, School of Biological
Sciences, University of Auckland, Private Bag, Auckland 92019, New
Zealand

FEATURES
source location/qualifiers
1..5284
/organism="Caldocellulosiraptor saccharolyticus"
/db_xref="taxon:44001"
279..341
misc_feature
/gene="celC"
/note="encodes leader sequence"
279..4374
gene
/gene="celC"
279..4374
CDS
/gene="celC"
/note="The celC gene potentially encodes a multidomain
enzyme comprised of an N-terminal endoglucanase domain,
two central cellulose-binding domains, and a C-terminal
mannanase domain. However, due to the presence of two
open-reading frame shifts in the mannanase-encoding region
of celC, the actual enzyme is truncated and comprises only
the endoglucanase and cellulose-binding domains"
/pseudo
/product="endoglucanase"
279..3572
misc_feature
/gene="celC"
/note="ORF #1 from celC start codon to frameshift #1"
342..1715
misc_feature
/gene="celC"
/note="encodes endoglucanase domain"
1722..2204
misc_feature
/gene="celC"
/note="encodes cellulose-binding domain (CBD) number 3"
2205..2342
misc_feature
/gene="celC"
/note="encodes proline-threonine linker number 1"
2343..2807
misc_feature
/gene="celC"
/note="encodes cellulose-binding domain (CBD) number 1"
2808..2957
misc_feature
/gene="celC"

misc_feature
/note="encodes proline-threonine linker number 2"
2958..3419
/gene="celC"
/note="encodes cellulose-binding domain (CBD) number 2"
3420..3524
misc_feature
/gene="celC"
/note="encodes partial mannanase domain"
3420..3524
misc_feature
/gene="celC"
/note="encodes proline-threonine linker number 3"
3574
misc_feature
/gene="celC"
/note="location of the first -1 frameshift"
3575..3625
misc_feature
/gene="celC"
/note="ORF #2 from frameshift #1 to frameshift #2"
3627
misc_feature
/gene="celC"
/note="location of the second -1 frameshift"
3628..4374
misc_feature
/gene="celC"
/note="ORF #3 from frameshift #2 to celC stop codon"
BASE COUNT 1829 a 881 c 1238 g 1336 t
ORIGIN

Query Match 8.5%; Score 125; DB 2; Length 5284;
Best Local Similarity 51.1%; Pred. No. 5e-19;
Matches 405; Conservative 0; Mismatches 370; Indels 18; Gaps 4;

QY 211 attgaagagatgcgaacaacgcggtcattacgcgcgagatggttctcgaatgggga 270
Db 3610 ATTGGTGGGAATCAGCGATGGGGATGAACTCTGTGAGGAGTGTGATGATGCTTAC 3665
QY 271 caatgacaagaatgacatcatcagtaagaacctctctttagcggaagataat 330
Db 3670 CGATGACGAAGATACCGACGACGATGAAGTGGCAATATATATATCTTTCAGAAAGCTT 3729
QY 331 cattgtgtcgtt--cttgaagtcattgattgctaccggtatgattcattg---- 382
Db 3730 GGTTCGAAGCTATATATATGAAAGTACACGACACAGATATGAGAGATGGGCGCA 3789
QY 383 ----cttgcctcaatcgcgctggtgattgattgattgaaagagagcttattgga 438
Db 3790 GCGTGTCTATTGTGACAGCAGTGAATATTGGAAGGATTAACAGGCTATTACCGT 503
QY 439 aaggaagatccgctcattatcatatgacgaatgattggttcgtgggaaggagat 438
Db 3850 AACGAAGATTTTGTATTTATTAACCTTGGCAATGAGCCGTATGGAAACAATTAATCTCA 3900
QY 499 gtttggctgcaggtatataacaagaacatcccgattgcgtaacgcccgtctaaacct 554
Db 3910 AACTGCGTTATGACACGACGACGATTAAGACACTTGAAGATGAGATTCACGAC 3963
QY 559 accttgatgtagatgctgcgggggtggagac--aattccaatcgattcattatcat 623
Db 3970 ACGATTAATGGTGGAGTCCCGCACTGGGTGAGATTGGTCTTAATCAATGATGATAT 4022
QY 616 ggaagaagaattttaatgctgcgaacctcaacgaatacattgattcattatcatg 675
Db 4030 GCCAGACATTAATGGAAGCAGATCCCTGCGCAATTTGGTATTTTCCATTTCATATGAT 4089
QY 676 gaatatgagtgtaatgctgcgaatgctgactaaattgaccgagttcttaataca 735
Db 4090 G--GTGATACAAATACGACGACGAGTGCAGAGTACATCAATCAATTTGTTGATAG 4146
QY 736 gacctcgcatgattgattggaatttgacacgctatatacaaatgagtgagtgatgaa 795
Db 4147 GGGTACCATTTGTTATTTGGGAAATTTGGACATCGACACAGATGACCGTGTGATCAA 4206
QY 796 gcaacgattatgactcttcgaaacaagagaggtgtggtgtggtggtggtgcatgaaa 855
Db 4207 GAAGCTATTGTCAGGTATGCAAAACAGTACAGATAGATATTATTAGTTGGTGGTGT 4266

[illegible]

Best Local Similarity 55.0%; Pred. No. 2.2e-15;
Matches 259; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

QY 1000 tctgtaggaaggaactctccgacacactcttattgtaagtagtatgacgaagtgtg 1059
DB 1288 TATGATCAATCAATCAATCAGCAGCTCTCTATGATTTTGAATAAATAGCATCTG 1347
QY 1060 actggaagtagctt---gagcggaagtccttggtgacgaagtagcttcttaagga 1116
DB 1348 TCCGGGTGTAATTTAGGAGGAGGACCATGACCTTCGATGATGATGCGGCAATAGGT 1407
QY 1117 agcatctcttaaaaggaagatattcatttcgtcaaatccacacacttaactatgctt 1176
DB 1408 ACTCAATCGTTGAAGAGATGCTGTTGGCAATATAGCTACAT---TTCGCAAAA 1464
QY 1177 attcaaatcagctcttaccagcagaatagtaggatacaagcactactgttaaacatgcaat 1236
DB 1465 ACAGTGAATCGAATCTTATTTCATTCAAAAACCTAGAAATTAAGTACCATCTTCG 1524
QY 1237 tggggaagtagtctgtaagtagatgactgcgcgtcttattgtgaagaacagacatggtat 1296
DB 1525 TGGGGAATGTAGGAGATGCGATGACAGCAAGAGTTTTCGTAACAGGAGAGCTTGG 1584
QY 1297 actgtgactctggaagcttctgacgataaaggttcacatcgtgaacacagctatctcta 1356
DB 1585 AGATGGAATCGAGTGAATTTGTCACTTTCAGAGCAACCAACGACGACTATCTATT 1644
QY 1357 gattatcaaatgctcaaatctcttccaaagtaaggaattgagttcagttccaatca 1416
DB 1645 GATTGACGAATGAGTAACTATCTGCATGATGTTGAGAGATAGTGTAGATATAAGCA 1704
QY 1417 gcgagtagatagtagtgacacacatcgattatattgataatgtagtattga 1467
DB 1705 CCAGCAATATGCAACGAGGAGGAGCGGATTTACTTATGATCATGAGACCGTA 1755

FEATURES

source

CDS are numbered using the following system eg SC2H4.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complement), strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in coding using a specially developed Hidden Markov Model (Krogh et al, 1994, Nucleic Acids Research, 22(12):4769-4778(1994)) and the EMBL program of Bibo et al., Gene 30:157-66(1984) as implemented at <http://www.nh.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted a correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid 2H4 lies between ZG5 and AH10 on the AseI-B genomic restriction fragment.
Location/Qualifiers
1..25970
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2H4"
1..1822
/gene="SC2H4.01"
1..1822
/gene="SC2H4.01"
/note="SC2H4.01", probable ATP/GTP binding protein, partial CDS, len: >606 aa; contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/transl_table=11
/product="putative ATP/GTP binding protein"
/protein_id="CA20595.1"
/db_xref="GI:3559957"
/translation="IDSTAPKAPVIRISLSPYOCNTADCEKGGPAGFTFOPNAA DDDINGYRRLITTSKKAAYTGKSVYSDTPSLSGOYLSVEAKDYRWGAPAE FFRKYPAGATGTHFDALPGSGVTTAKDTAABTRHADLTLYTGAGWSTLARGD AONSLWLDNGSDTOTRAGYAATSAFVNRDSFTVSAMAVLTDTSOTRVMAAPGTG SAFETLYSASRYKRWENRAGVNDKPYRLRSIGDAPLPVTHLAIVPRTKDTN KANDITQLFVNGRPGEPYLDGVSAFVPGVSGGLQIGSLVGGAMEHFRGRIDE VNMVORVLPDELTQOALLEGVAPNELVAMQMDAASGTTGKELSSVPASLTLSA ACAYVDENNALVILDSAGYASATGVDNTESPFYSAPVQIDSLAKSPKYEVLV AACQASAGSSNALNAYKPADDEVYOKRTTRIVAGADKVTQSAEAPGCIATETVQ VTGVGDAQEPWEMETPDADSAKTEIRYGRILVYGERPDQSTFPTQYAGALAV GRSRGCTTGTHLPALIELRWVTGMSADQVRSOVLGCV"
188..211
/gene="SC2H4.01"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
1949..1953
/note="possible RBS upstream of SC2H4.02"
1959..8510
/gene="SC2H4.02"
1959..8510
/gene="SC2H4.02"
/note="SC2H4.02, unknown, len: 2183; very limited similarity to proteins of the RBS family e.g. WAPA_BACSU wall-associated protein precursor (2334 aa), fasta scores: opt: 141 z-score: 296.8 E(1): 2.7e-09, 23.4% identity in 1237 aa overlap, and RBSA_ECOLI RBSA protein precursor (1377 aa), fasta scores: opt: 135 z-score: 222.4 E(1): 3.8e-05, 23.0% identity in 1016 aa overlap. Contains

RESULT 9
SC2H4 25970 bp DNA BCT 19-OCT-1999
LOCUS Streptomyces coelicolor cosmid 2H4.
DEFINITION AL031514
ACCESSION AL031514.1 GI:3559956
VERSION
KEYWORDS aminotransferase; beta-mannosidase; permease; secreted beta-mannosidase; sugar transport system; sugar-binding lipoprotein.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
REFERENCE 1 (bases 1 to 25970).
AUTHORS Redenbach, M., Kleser, H.M., Denapate, D., Elchner, A., Cullum, J., Kinsahl, H. and Hopwood, D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
REFERENCE 2 (bases 1 to 25970).
AUTHORS Oliver, K. and Harris, D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 25970).
AUTHORS Parkhill, J., Barrett, B.G. and Rajandream, M.A.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmid supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)

Db 25173 ACGGCCAATCCCTCGGCTCGCTACCTCGCTGCTCTGAGCGGACGCGCGGC 25232
QY 871 tggagatattagacccttcgaatgaltggtggaataacctaagcttggggaat 930
Db 25233 GTCGAGTACTGACACTGTGCAACGGCTTCGACCGCAACTCGCTGACGAGCGGGCAAC 25292
QY 931 acaatagtgatgtccatatagttttaagagaacttcgagacttaagcagcttttaca 990
Db 25293 CCGACTCTTACGAGGAGCAACGGCATCCGCCGACGCTCCAGGACGCGCATGTACGCGC 25352
QY 991 ggtgagag 998
Db 25353 GGTGGTGG 25360

RESULT 10
STMBMANASE 1881 bp DNA BCT 16-JUN-1999
LOCUS Streptomyces lividans mannanase (mana) gene, complete cds.
DEFINITION M92297
VERSION M92297.2 GI:4579683
KEYWORDS
SOURCE Streptomyces lividans.
ORGANISM Streptomyces lividans.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS Arcand, N., Kluepfel, D., Paradis, F., Morosoli, R. and Shareck, F.
TITLE Beta-mannanase of Streptomyces lividans 66: cloning and DNA
JOURNAL sequence of the mana gene and characterization of the enzyme
MEDLINE Biochem. J. 290, 857-863 (1993)
93207541
REFERENCE
AUTHORS Shareck, F.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1992) Centre de recherche en microbiologie
Appliquee, Institut Armand-Frappier, 531 boul. des Prairies, Laval,
Quebec H7N 4Z3, Canada
3 (bases 1 to 1881)
REFERENCE
AUTHORS Shareck, F.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1999) Centre de recherche en microbiologie
Appliquee, Institut Armand-Frappier, 531 boul. des Prairies, Laval,
Quebec H7N 4Z3, Canada
Sequence update by submitter
On Apr 8, 1999 this sequence version replaced gi:153193.
REMARK
COMMENT Location/Qualifiers
FEATURES
Source
1..1881
/organism="Streptomyces lividans"
/strain="1326"
/db_xref="taxon:1916"
217..1316
/gene="mana"
217..221
/gene="mana"
/note="putative"
226..1377
/EC_number="3.2.1.78"
/codon_start=1
/transl_table=11
/product="mannanase"
/protein_id="AAA26710.2"
/db_xref="GI:4579684"
/translation="MRNARSLITITAGNAPVLIHLLFALAGPSAGRAEAAGIIVSN
GRVYENGSAFVGMGVNNAATYMPDRGTSIDIAKAGANTVRYVLSGGRATKTSASE
VSALIGCKAKKAVICVLEVDHTTGKRGATSLDAGTYWGVSAARAEEDTYVNA
IGNEPFGTNTAAATDARKSAIGLRGAGLHALLVNDAPNNGQDSGTRSNASVFA
SDPDNNTVFS.IHMGVYDPAEVRDYLAFVNGNLPYVGEFGDHSNGNDEDAIMA
TAOSIAGVYLGMSWNGSGNGVYLDNMGFDPNSLTSGNRLIYGSNLIATSRATV
YGGGGGTGTGATPANGYPCVNGASDPDGDGMGWNSSCVVRGSAADH"
226..333
sig_peptide

mat_peptide /gene="mana"
334..1374
/EC_number="3.2.1.78"
/function="hydrolase"
/product="mannanase"
BASE COUNT 309 a 648 c 647 g 277 t
ORIGIN

Query Match 6.1%; Score 89.2; DB 74; Length 1881;
Best Local Similarity 47.6%; Pred. No. 1.2e-10;
Matches 432; Conservative 0; Mismatches 458; Indels 18; Gaps 5;

Db 103 ggaattatgtatgagcgttacccatctatagcgtgcaatggaaccattgtaagaga 162
Db 340 GGCATCCACGTCAGCAACGAGCGAGTGTGCGAGGAGCAACGCGCGCTTCATGCGCC 399
QY 163 ggaattaaacatgggacagcatggtatagaagccagcgaactcgaattgaaggatt 222
Db 400 GGGGTCAACCGCTTACCTGATCTGTTCCGAC--GCACCGCTTCATGCGCACATC 456
QY 223 gcaataccggtgctaatcagtcggtatggtatctgatatggtgggagacaatgaca 282
Db 457 GCGGCCAAGGAGCAACACCGCTCCGCGTGTGTCAGCAGCGCGCGCTGACGAGAG 516
QY 283 gatgacatccatcacgtaagaacctatctctttagcggagaataatcattggtgct 342
Db 517 ACGACGCGCTCCGAGATCTCGCGCTCATCGGCCAGTGCAGAGCCACAGGTCATCTGT 576
QY 343 gttcttgatgacatgacgtcctcggtatgattcattgctgctca-----atgct 396
Db 577 GTCCTTGAGTGCACGACACCCAGCGCTACGGCAAGAGAGCGGCGCACCGCTGAGCAG 636
QY 397 gctgtgattatgattgaaatgagaagtgct---taattggaaggagaatacagtc 453
Db 637 GCGGCGACTGAGTGGGTGCTGTGAGAGCGCGGCTGAGAGGCCACGAGAGACTACGTC 696
QY 454 attatataatgtcgaatgaatgattggttcgtggaaggagtgcttggtgacggg 513
Db 697 GTCGTCAACATGCGCAACGAGAACCTTCGCGCACACCACTACGCCGCGGACCGCG 756
QY 514 tataacaagcatcccgatggtgtgtaacgggtctaaacatacattggtgtagat 573
Db 757 ACCAAGTGGCGCATGCGCAAGCTCCGCGCGCGGCTGCGCACGCCCTGATGTCAC 627
QY 574 gctgctgaggtgagac---aatccacaatcgattcatgattatggaagagaagtttt 630
Db 817 GCGGCCAATGAGGGGCGCAGAGACTGTCGCGACAGATGCGGTCCAAAGCGCGCTTC 876
QY 631 aatgtgacctcaacgaatacaatgatttcgattcatatgatatgaaatgcaagtggt 690
Db 877 GCGTCCGACCCGACCGACACCGCTTCCTTCATCCACATATAGC---GGGTCTACGAC 933
QY 691 aatgcatgcaagtcgtaactaatatgtacgaggtcttaatacaagactcgtcatagtc 750
Db 934 ACCGCGCGCGAGGTGCGGACTACCTGAACGCTTCGTGCGCAAGAGATGCGCATCGTC 993
QY 751 attgtgaatttgcacacgcatatacaaatgtgtgacgttcgtaaggaagcaagattagc 810
Db 994 GTCGGCGAGTGGGAGCAGACAGTAGACGCCAACCCGAGAGAGAGCGCATCATGGG 1053
QY 811 tattctgaacaaagagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 873
Db 1054 ACGGCCAAGTCCCTCGGCGTCCGCTACCTCGGCTGCTGTGAGCGGCAACGCGCGCGC 1113
QY 871 tggagatttttagacccttcgaatgaltggtggaataacctaagcttggggaat 930
Db 1114 GTCGAGTACTGACACTGTGCAACGGCTTCGACCGCAACTCGCTGACGAGCGGGCAAC 1173
QY 931 acaatagtgatgtccatatagttttaagagaacttcgagacttaagcagcttttaca 990
Db 1174 CCGACTCTTACGAGGAGCAACGGCATCCGCCGACGCTCCAGGACGCGCATGTACGCGC 1233

JOURNAL MEDLINE 97000351
 REFERENCE 2 (bases 1 to 14922)
 AUTHORS Seeger, K. J. and Harris, D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 14922)
 AUTHORS Parkhill, J., Barrell, B. G. and Rajandream, M. A.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
 Streptomyces coelicolor sequencing at the Sanger Centre is funded by the BBSRC and Beowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
 CDS are numbered using the following system eg SCF7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh.90.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
 Cosmid F73 overlaps with cosmid F11 on the AseI-F genomic restriction fragment.

FEATURES
 source 1. 14922
 location/Qualifiers
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid F73"
 1. 122
 /note="nominal overlap with str11"
 complement(1..983)
 /gene="manA"
 complement(<1..983)
 /gene="manA"
 /note="SCF73.01c, manA, probable secreted beta-mannosidase, len: 327 aa; similar to many e.g. MANA, STRLI P51529 mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189 z-score: 1332.7 E(): 0, 61.1% identity in 288 aa overlap. Contains N-terminal signal sequence, and Pfam match to entry PF00150 cellulase, Cellulase (glycosyl hydrolase family 5). Also similar to SC2H4.16 (62.2% identity in 288 aa overlap)."
 /codon_start=1
 /transl_table=11
 /product="putative secreted beta-mannosidase"
 /protein_id="CAB57406.1"
 /db_xref="GI:6013077"
 /translation="MRPARRDTPTPLRLTGLALIGLVVVGALCPGALASPAA

RSAVPSSAGQAAGLHIGDGRLEBNGNDPFMRGVNNAHMTVPGETOSIADYKALGAN
 SVRYVLSGHRHSENGRPAADVAAYIEOCKANLICEVLEPHDTGAEADRACTLDHAAD
 TWIGKIDVLAGEDEYIVNIGNEPMTNDPAGMTEPYVAVKRLAAGLOHTIVADP
 NMGDQGVKRNARASVYADNPOTNLIFSIMYSEVFTQAEITDYLAAPDAEPIILI
 GEFGPPADQYGDPEDETMATAEQLRLGILAMSSMGNDPVLIALDLPDS"
 complement(989..992)
 complement(1129..2724)
 /gene="SCF73.02c"
 complement(1129..2724)
 /gene="SCF73.02c"
 complement(1129..2724)
 /note="SCF73.02c, possible membrane-bound oxidoreductase, len: 531 aa; some similarity to eukaryotic molybdenum-containing oxidoreductases e.g. SUOX from Sulfolobus solfataricus (488 aa), fasta scores: opt: 1189 z-score: 281.9 E(): 2.5e-08, 30.4% identity in 319 aa overlap, and NIAL_ORYSA nitrate reductase 1 (916 aa), fasta scores: opt: 247 z-score: 262.1 E(): 3.1e-07, 28.9% identity in 336 aa overlap. Contains several membrane spanning hydrophobic domains near N-terminus, and binds molybdopterin binding domain"
 /codon_start=1
 /transl_table=11
 /product="putative membrane-bound oxidoreductase"
 /protein_id="CAB57407.1"
 /db_xref="GI:6013078"
 /translation="MSDDKTPRNPSPARTLRGALSGILLAGALAVELAAVR
 PRSGPVAAGGAAIDRPTAVKDAIRFGINDVLQGLAVLTFALALGFAVR
 HRRAGAGVLEFGVGAALAGRDSSTGTFAPSGAIGAVLALDLALDLPK
 EAPERGGMRRRFRVATATAAASGTGVLGRALSGAGREAVASENTLEPPS
 RAPVPRQARVSGVSTPTPTGDFYRDVATLVPPADATMLRIHGEVARENL
 TFDLLRLKRLERITITLCVSNKGVGPAKRIQVRLADLLCEGRPSAGAC
 QLVARSVDGMTGSPVEDVMDGRUMALVANGSEPLPDHGFVRAVVPGLYGVSA
 KMIDIELTFPDSDAYVKKDMARERAPVKQSIDTPRPARAKGVAVAGAAWQ
 HRGIDKVEYRDDPMEAEVLAADSBDTMRQMSVAMRATKGTTLVRADRGVWQ
 TDKRTRVYDCAGRSHSVYVVD"
 complement(12777..4300)
 /gene="SCF73.03c"
 complement(12777..4300)
 /gene="SCF73.03c"
 complement(12777..4300)
 /note="SCF73.03c, unknown, len: 507 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCF73.03c"
 /protein_id="CAB57408.1"
 /db_xref="GI:6013079"
 /translation="MTAFILVSGMFTGTHIMQDPAARLTARGAEVHTALGLDAPRA
 AARGVDLETHIADVLAVDSVGAAGDRIVLVGHDIHRAVGAADRRARVRIVH
 IDSGLPDRGVFALAAVDPQSLRELRAGAAGDYLPPRAHEMRKSTAGPVAALD
 RLIALAQPGLSTLQPLRLTGAVDPVPTGVLCTGNTSTELVQMLRLGDPALRL
 TDPVSEFEELPTGHPMLSLPAELTDLVLRAAAGRLEVPDVTAGGHLRPLMDV
 PDVPRERGNLDLVPDAEPRPAVVVHGGPVPADARPTPRDMPGLTGYARCYAGD
 AVGALDLRLHLDGDFERAAADVAEAAGVADRDVDAALWFEISGGGILADMDL
 APAMRLICATVYPLAVLPWAGVSESRFVRVAVANGALPVLLTRGRMPELAT
 VERLLAAKCGAGVEVVDVPHGHGFTIDLTDESRAVNAHRLRYLDHAFGHAR"
 complement(4311..4315)
 /gene="SCF73.04"
 4382..4387
 /gene="SCF73.04"
 4392..5414
 /gene="SCF73.04"
 4392..5414
 /note="SCF73.04, probable transcriptional regulator, len: 340 aa; similar in N-terminus to members of the MerR family e.g. NOLA_BRSN nodulation protein NOLA (237 aa), fasta scores: opt: 211 z-score: 233.9 E(): 1.2e-05, 48.5% identity in 66 aa overlap. Also similar to SC017.06c (34.8% identity in 322 aa overlap). Contains Pfam match to entry PF00376 merR, Bacterial regulatory proteins, merR family"
 /codon_start=1
 /transl_table=11
 /product="putative MerR-family transcriptional regulator"
 /protein_id="CAB57409.1"

/db_xref="GI:6013080"
/translation="MTSNDGLGICIGLAGAGVYVYTFVPSDRGLIPASRSAGH
RYPGPEALDRRLIRSLRALGLVPEVRLILDEHNAHTGGSGALAEAVGRLEEV
GSELRALMREALRLVQEAPEGOADRLRLIGAVSPSTALVFRKMLPARRPA
RSAGFLDVAVPDPDEPAPAPVLAFLNALTLAPRPTVPOPEFAHAAARAPA
LLYGLAEVLAAGVHMRGRPEGLALAFVASYSAVNDTPEFRRLAQLAA
DPRIDGVELVAEYTTAPGARPEPTPSADHMLAALTAGTAVPDGTSPROSKRRV
SPGLGV"
gene complement(5357, .5623)
/gene="SCF73.05c"
CDS complement(5357, .5623)
/gene="SCF73.05c"
/note="SCF73.05c, unknown, len: 88 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.05c"
/protein_id="CAB57410.1"
/db_xref="GI:6013081"
/translation="MDEVFMRGRVYGDHNPGRAGRVYAHVVGEPDGLLDVTD
LTPDRARGVTLATETIGRYPGSHSATYPRRGTDRFRDMDVP"
gene complement(5696, .6697)
/gene="SCF73.06c"
CDS complement(5696, .6697)
/gene="SCF73.06c"
/note="SCF73.06c, unknown, len: 333 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.06c"
/protein_id="CAB57411.1"
/db_xref="GI:6013082"
/translation="MLNDTPVYVDRGRYDYTPMEEPARTEATGMEETARLAAGLR
PTGRMVRLEPMSVILRLCVAAGDAVWFASPPAGLEALASTARVPRVLEPLA
VDARMSILPDPDWELFVLAEPDPAWELPDLQYAMHALTSRSETERLAP
AVAGAGVAGLIDLPALSRDALROLRPLDMLCETLALGVPSLDIADHDG
QVPRPAPGRTPEMDGDAVYTOPFSSILAVPARAAVYPOVLPRLRDVLEPMTGEG
RTAQLEKRAVSLMRLSALNRAAYGRLLPFGASGAGATTAAGANCLLELDEPPL"
gene complement(6778, .9000)
/gene="cpeb"

Query Match 3.6%; Score 52.4; DB 2; Length 14922;
Best Local Similarity 46.9%; Pred. No. 0.045;
Matches 275; Conservative 0; Mismatches 296; Indels 15; Gaps 3;

QY 101 cggatttattgtaagcgggtaccactctatcgagcaatggaacccattgtaata 160
DB 814 CCGGCTGCACATCGGTACGCGCCGCTGCGAGGCAAGCGACACTCTCTATGC 755
QY 161 gaggaattacattgagcagcgtatgaagaccagcactactgcaattgaagga 220
DB 754 GCGGGGTCAACCGCCACCTCTGATCCGCGGAGAACCGCTGGCGGACGTCA 695
QY 221 ttgcaataccggtgtaatacgcgtcgatgtgtatctgtagggggaacatgggaa 280
DB 694 AGGC--GCTGGCGGGAACGCTCGGAGTCTCTCCACGCGGACCGCTGGAGCG 638
QY 281 aagatgaacatcatagcaagaacactatctcttagcgggaagatatacttggty 340
DB 637 AGAACGCGCGCGGACGTCGCCCGCTATCGACGAGTGCAGGCGCAACCGGCTCATCT 578
QY 341 cgttcttgagttcgtatgtaacggtatgattcatcgtc-----ttgcgtca 391
DB 577 GCGTACTGAGGTGACGACGACACCGGCTACGCGGAGACCGCGCGGACGCTCG 518
QY 392 atcgtcgtgttatttggatgtaagatgaagatgcttaattggaaggaataccg 451
DB 517 ACCACGCGCGCGGACTACTGATGCGCTCAAGGACGCTCTGGCGCGGCGGAGACTACG 458
QY 452 tctttttaaattgcaatggaatggttggctgtaggaagggagatgcttggctgacg 511
DB 457 TCATCGTCAACATCGGAGAGACCGCTCGGCGCAACCGCGCGGCTGGAGCGAGC 398
QY 512 ggtataaacaagaatcccgatgtgtagtaacgggtcttaaacatacttgatgtgtg 571

DB 397 CGACCGTCCGCGCGCTCAAGAGCTGCGCGCGCGCGGCTCCACACAGATCATGTGCG 338
QY 572 atgctgcyggggtggggaacattccacaatg--atcagattatggaagagaagttt 628
DB 337 ACCGCCCCAACCGGGCGCGAGCTGCGAGCGCGGTATCTGTCACACGACGTCCTCT 278
QY 629 ttatgtctgacctcaagaataacatggttttgcattcatatgta 674
DB 277 ACGACGCGGACCGCGGACCGGCAACCTGATCTTCTCATCATCATGTA 232

RESULT 13
LOCUS 166494 7218 bp DNA PAT 28-DEC-1987
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
LOCATION/Qualifiers
SOURCE 1. 7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 3.3%; Score 48.6; DB 5; Length 7218;
Best Local Similarity 3.8%; Pred. No. 0.35;
Matches 15; Conservative 216; Mismatches 160; Indels 0; Gaps 0;

QY 747 agtcattgtgaatttggaacccgtctatcaacaatggtgagctgtagaagaacagattat 806
DB 1455 ACAGATAGACAAATTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1396
QY 807 gggcatcttgaaagaagagatggtggtgctggtcatggaaggaagggcc 866
DB 1395 RRR 1336
QY 867 agaatgtagatttggaaccttgcaatgattgggctggaataacttaagcttgagg 926
DB 1335 RRR 1276
QY 927 aatacaatagatggtgcctatggtttaagaagaacttcgaataacttaagcaccgttt 986
DB 1275 RRR 1216
QY 987 tacagttagagatctgtagagaactctccgaacaactctttagatttgaagtag 1046
DB 1215 RRR 1156
QY 1047 tatgcaagatgtagcctgtagatgcttgtagcggaggtccttggtcgtgtagaagtg 1106
DB 1155 RRR 1096
QY 1107 ttctaaggaagtcattctttaaagcggat 1137
DB 1095 RRR 1065

RESULT 14
AC024591 111071 bp DNA HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 16 clone RP11-511G21, WORKING DRAFT
DEFINITION
SEQUENCE, 13 ordered pieces.
ACCESSION AC024591
VERSION AC024591.3 GI:9256450
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (28-OCT-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 28, 1998 this sequence version replaced g1:1280123.
Submitted by:

Genome Sequencing Center,
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rwenemate@wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is C04D1, 2700 bp overlap; 3' clone is F49E7, 200 bp overlap. Actual start of this clone is at base position 701 of CELK03E6; actual end is at 35049 of CELK03E6

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source

1. 35049

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/clone="K03E6"

/chromosome="X"

/complement(3691..7748)

/gene="unc-1"

complement(join(3691..3766,3931..4145,4822..5095,5154..5278,6432..6545,7296..7326,7735..7748))

/gene="unc-1"

/note="K03E6.5"

/codon_start=1

/product="Erythrocyte band 7 integral membrane protein"

/protein_id="AAC6904.1"

/db_xref="GI:1280125"

/translation="MEYSHQIPKSCATIDVPPDYETIGTIFGVALQMSLIIIVF
PMSVCVLIVKEVERVIFRIGLVFGARPGMIFIPIDYRKIDLVSTAVP
BOETISKDVSVDAVYVFRSDPIASVNNDDIVSTKLAOTTLNAGMTVE
MTERAIAIOLCETITDECTEHGKVEVEVKDRLCOQTRAAVAEAARARAK
VVAABEQKASRAIKRAADYIQAIVALQRLQALNSIAEHNSTIYFVPVMEFGA
FMKDKQ"

complement(12649..14893)

/gene="K03E6.4"

complement(join(12649..12750,12801..12928,13397..13527,13570..13666,14217..14391,14837..14893))

/gene="K03E6.4"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC6904.1"

/db_xref="GI:1280126"

/translation="MERADSVILIHGNVTFEPESNVCKLSTVATGPKIINGE
TVLSFSFRGADNOIFIDPEFGKGSFPMPLVNCFSKDSKDCVGTIGML
PIPEPEKHYCHVCFLEPESSTVOOLIAIRVNAFVDPMLPANAADRYACRTSTK
GYVDLEINRITKCSDMARKFTSTSTKSSNSHATNNVSTTIADLAGIENVNEETFN
AOEDNEESKR"

gene

CDS

20796..24499
/gene="lim-6"
join(20796..20881,21727..21835,22141..22262,23465..23885,24151,24209..24499)
/gene="lim-6"
/note="K03E6.1"

/codon_start=1
/product="similar to other homeobox domains; contains two LIM domains"

/protein_id="AAC6904.1"

/db_xref="GI:1280127"

/translation="MSLLISATYSSTEDKLGSCGCLIKOREFLIRIRFSCQMA
HILVDFITITVLDKSNMDESYHESCLRSCQSLSSFKCSRGNTICCH
DHOMVAGKRCRCMTLLPTDVRHVMHYHAOCFSCSCORFNDDEYHVDGEV
FCNDKQSCINQOTSEPHYTLFTTFPAISVPDIMEEVRSERIKRKLKSDSG
AQRORFKTAFERSKSPSKRYEOLANETGVSRYVYVWFONORAKIKLKDSDG
DFKFGSGSGSTEDIRSDDEBSVLSKLRIQIDIGELMLYKLSVQKRVFVP
"

gene

CDS

complement(25406..26485)
/gene="K03E6.6"
complement(join(25406..25503,25556..25679,25726..25773,26375..26485))
/gene="K03E6.6"

/note="Similar to Profilin; K03E6.6"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC6904.1"

/db_xref="GI:1280128"

/translation="MSMSDIINNLIIGSNVSKAAILGFGAVMAKSDNFNISVEAV
/AAGAFSLDALGTCIRLEQKEFLVNDNDILIKOGSGSEFIYKTIQAVIISYE
KGLQPEMCSRTGALADYRSIKY"

complement(28061..33962)

/gene="K03E6.7"

complement(join(28061..28224,28960..29077,30844..30997,31048..31205,31305..31629,33214..33247,33211..33354,33555..33654,33708..33861,33914..33962))

/gene="K03E6.7"

/note="Similarity to C. elegans protein C01C10.4"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC6904.1"

/db_xref="GI:1280129"

/translation="MDETATSEVTEFVSDDPTTRPFEDGHHPLETRHLMTHELE
KLINSTDVNNMEYDOLARADPEROVONSEKLEKSYSSQIAKAPFYELKIKE
BSURESKAERBERBATSTIGIAKOVOVLOSSTVLEPCLEVLNHHIOYRE
VEERTTAESLHASNAHMLAKIKRAMENDNAYIKRSLEYEKLEPTKILEAK
ATILCLEAEVQKNDYTSUNLZERISERIERHRSRGSLEAVSPOEDOKSPKES
ESIPENPPVAPVAPPPEDYIIDKDDSVLNMRTEDQEBEGRNSRSLSGVIL
LAOOLIGNGNSTEKNITPPRHGEADISYTRVGVSDGSDNSEVSLASFNIGDIL
TVSKMLMSHELKCECAIKTEAIVAVAKFONRDLPKQVADNRRLRPITPCKQ
IKNSITLIDPFLANOSIVFESMAVGRGYPFPLSLCEFRVSGTGR"

BASE COUNT
ORIGIN

11644 a 5816 c 6271 g 11318 t

Query Match 3.1%; Score 45.2; DB 33; Length 35049;

Best Local Similarity 51.6%; Pred. No. 2.1; Mismatches 118; Indels 2; Gaps 1;

Matches 128; Conservative 0;

196 caagcaactacgcaatgaaggatgcaataacccggtgcaatacaggtccggatgtg 253

Db 33790 CATGAATCATATTCAAATGACGAGTTCCAGTGGAGGGGTGCATCTTTCGAATTCG 33810

Qy 256 ttatctgatggggacacatgacaaagaatgacatcacatcagtaagaacattatctt 315

Db 33850 TTGTTGGGGGTCTCGAATTCAAAATTCGAAATTTAGAAATTAAGAAATAGATGCT 33900

Qy 316 ttacggaagaataatcttggttcgtctcctgaagttcatgattaccggttatga 374

Db 33910 TTACTGAAATAAATGTTTCGTGACCTCCGAAAGAGTTTCAGATTTCGTCATTTTGTA 33969

Qy 375 -ttcatgttcgtcgaatgctgctgttatgattgattgaatgagaatgctttaa 423

Db 33910 TTACTGAAATAAATGTTTCGTGACCTCCGAAAGAGTTTCAGATTTCGTCATTTTGTA 33969

Qy 375 -ttcatgttcgtcgaatgctgctgttatgattgattgaatgagaatgctttaa 423

Db 33970 CTTTATAGGTCATGCAATTTTGTGCGTGTCTTTGTTCTGTGAGTAACATTTTAT 34029

OY 434 ttggaag 441

|||||

Db 34030 GCGGAATG 34037

Search completed: December 20, 2000, 03:12:31
Job time: 12796 sec

This Page Blank (uspio)


```

DB 1261 acggcgcttcttatgtgaaacacgacacggtatcatatcgttctcttggaagcttctg 1320
QY 1321 ccgattaacggttcatcttggaacacgctatctcttagattatcaaatgccaacatctt 1380
DB 1321 ccgattaacggttcatcttggaacacgctatctcttagattatcaaatgccaacatctt 1380
QY 1381 tctcaagtaagggaatgtgagttcagttccaatcagcgagtgtagtagtgagcaaca 1440
DB 1381 tctcaagtaagggaatgtgagttcagttccaatcagcgagtgtagtagtgagcaaca 1440
QY 1441 ttcgatttatgtatgtatgtatgtatgaa 1470
DB 1441 ttcgatttatgtatgtatgtatgtatgaa 1470

RESULT 2
ID 245336 standard: DNA; 1438 BP.
XX 245336;
AC 245336;
XX 27-MAR-2000 (first entry)
DE DNA encoding a mannanase-linker-cellulose binding domain fusion.
XX
XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX endo-1,6-mannanase; Bacillus sp. 1633; galactomannan;
XX 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
XX galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
XX printing paste; plant material degradation; recycled waste paper;
XX paper making pulp; guar; locust bean gum; thickener; viscosity;
XX mannan-containing food; coffee extract; cleaning composition;
XX machine washing; hard-surface cleaner; dishwashing; oral; dental;
XX contact lens; body-care composition; fabric softener; oil well drilling;
XX subterranean formation fracture; cellulose binding domain; ss.
XX
OS Synthetic.
OS Bacillus sp.
OS Clostridium thermocellum.
XX
XX Key Location/Qualifiers
XX CDS 1..1431
XX FT /*tag- a
XX
XX WO964619-A2.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1999; 99MO-DK00314.
XX
XX 10-JUN-1998; 98US-0111256.
XX 20-OCT-1998; 98DK-0001340.
XX 20-OCT-1998; 98DK-0001341.
XX 28-OCT-1998; 98US-0105970.
XX 28-OCT-1998; 98US-0106054.
XX 23-DEC-1998; 98DK-0001125.
XX 05-MAR-1999; 99DK-0000306.
XX 05-MAR-1999; 99DK-0000307.
XX 05-MAR-1999; 99DK-0000308.
XX 05-MAR-1999; 99DK-0000309.
XX 09-MAR-1999; 99US-0123543.
XX 10-MAR-1999; 99US-0123623.
XX 10-MAR-1999; 99US-0123641.
XX 11-MAR-1999; 99US-0123642.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX WPI: 2000-105891/09.
XX DR P-PDB; Y54123.
XX

```

```

PT New mannanases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions
XX
XX Example 4: Page 210; 242pp; English.
XX
XX The present sequence encodes a mannanase-linker-cellulose binding
XX domain fusion protein. Mannanase (also known as mannan
XX endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
XX hydrolyses galactomannans. Specifically, mannanases hydrolyse
XX 1,4-beta-D-mannosidic linkages in mannan, galactomannans, glucomannans,
XX and galactoglucomannans. The mannanase protein, or preparations
XX containing it, are used to improve properties of cellulosic or
XX synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannanases
XX are also used to process coffee extracts (to inhibit gel formation);
XX in cleaning compositions (for machine washing of fabrics, as
XX hard-surface cleaners, for hand or machine dishwashing, also in oral,
XX dental, contact lens or body-care compositions) where they remove
XX mannan-containing soils and prevent binding of some soils to
XX cellulosics; and in fabric softeners. They can also be used in oil
XX well drilling to fracture subterranean formations.
XX
XX Sequence 1438 BP; 445 A; 267 C; 344 G; 382 T; 0 other;
SQ
XX
XX Query Match 61.0%; Score 897.2; DB 21; Length 1438;
XX Best Local Similarity 99.7%; Pred. No. 3,8e-241;
XX Matches 899; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 94 gcaaatccgagatttatgtatgaacggtacacactctatagatgccaatggaacccatt 153
XX DB 1 gcaatccgagatttatgtatgaacggtacacactctatagatgccaatggaacccatt 153
XX
XX QY 154 gtaatggaaggatttaacccttggcagcgtatgtatataaagcaggaaactactgcaac 223
XX DB 61 gtaatggaaggatttaacccttggcagcgtatgtatataaagcaggaaactactgcaac 223
XX
XX QY 214 gaaggagattgcaaataccggtgtcgaatacgttcgaattgttatctgtatggggagca 273
XX DB 121 gaaggagattgcaaataccggtgtcgaatacgttcgaattgttatctgtatggggagca 180
XX
XX QY 274 tggacaaagatgacatcacatagtaagaaccttactcttttagcggaagataatcat 333
XX DB 181 tggacaaagatgacatcacatagtaagaaccttactcttttagcggaagataatcat 240
XX
XX QY 334 ttggttcgttcttgaagtcatatgatacgcggtatgatccatgttgcctcaat 393
XX DB 241 ttggttcgttcttgaagtcatatgatacgcggtatgatccatgttgcctcaat 300
XX
XX QY 394 cgtgccttgattatgattggaatgagaagtgcttaattggaaggaagataccgtc 453
XX DB 301 cgtgccttgattatgattggaatgagaagtgcttaattggaaggaagataccgtc 360
XX
XX QY 454 attactaatactgcgaatgagttgtgttcgttgggaaggagatgcttggcgtaacgg 513
XX DB 361 attactaatactgcgaatgagttgtgttcgttgggaaggagatgcttggcgtaacgg 420
XX
XX QY 514 tataaaagaagaatcccgagattgcgtaaagccggtctaaacataacttgatggtat 573
XX DB 421 tataaaagaagaatcccgagattgcgtaaagccggtctaaacataacttgatggtat 480
XX
XX QY 574 gctgcggggtggggaacaattccacaatgcatcatgattggaagagaagcttttaac 633
XX DB 481 gctgcggggtggggaacaattccacaatgcatcatgattggaagagaagcttttaac 540
XX
XX QY 634 gctgacctcaacgaataatacaatggttttcgattcatatgatatgaaatgaggttgtaat 693
XX DB 541 gctgacctcaacgaataatacaatggttttcgattcatatgatatgaaatgaggttgtaat 600
XX
XX QY 694 gcatcgcaagtcgtactataatattgacccgagttcctaatacgaacctcgcatgacatt 753

```


PR 10-MAR-1999: 99US-0123623.
 PR 10-MAR-1999: 99US-0123641.
 PR 11-MAR-1999: 99US-0123642.
 XX (NOVO) NOVO-NORDISK AS.
 PI Kaupinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX WPI: 2000-105891/09.
 DR F-PDB: Y54125.
 XX
 XX New mannanses for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 PT
 XX
 PS Example 5; Page 214-215; 242pp; English.
 XX
 XX The present sequence encodes a *Bacillus* mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The present mannanase is a synthetic variant of the mannanase of
 CC Y54124, in which the C-terminus of the protein was changed due to
 CC design of a lower PCR primer used for amplification. The mannanase
 CC hydrolyses galactomannans. Specifically, mannanses hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanses
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 CC
 CC Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other;
 XX

Query Match 41.4%; Score 608.2; DB 21; Length 1407;
 Best Local Similarity 65.6%; Pred. No. 2,2e-160;
 Matches 921; Conservative 0; Mismatches 473; Indels 9; Gaps 2;

QY	422	gaagtccttaattggaagaagataccgctcatttaatttaattgcgaatgagtttg	481
DB	422	aagatgcttattatcgttaagaagatacgtgtattatataacattgcaaaagatgtatg	481
QY	482	gttcgtggaaggagatgctgctgaacgggtataacaagaacatccgcgattcgta	541
DB	482	ggagttggatgctgcctgctggcgatggtctatatgtatgcatccgaagcttcgcg	541
QY	542	agccggtctaaacataccttgatgtatgctgcgggtggtggacaattcccaaat	601
DB	542	atgcgggtataacacacacctaattgttgcagcagagatggggacaatccgcgaat	601
QY	602	cgattcatgattggaagaagatttttaattgtgacccccaacgaataacattgttt	661
DB	602	ctatcagaattacggaagaagatgttttaattgcatccgcgttaaaaaataacgttct	661
QY	662	cgattcatatgataatgataatgcaagttgtaattgcatccgcaattcattacc	721
DB	662	ccatccatattatgagatgtgtgtgtatgtaacacacgttagataataattgata	721
QY	722	gagttccttaacagacctgcattagtcattgttgaatttggacaccgltacaaatg	781
DB	722	gagttcatatgaagaccttgcctcgttaaaaggtgaattgcgcacagacatactatg	781
QY	782	gtgacgtcagatgaagcaacgattatgactattctgacaagaaggagttggtgttg	841
DB	782	gtgagttgtagaagaagatacattccatttcttgcgaagaaccggaagggtgtctg	841
QY	842	cggtgcatggaagaagggaagcccaaatggtggatttgaaccttcgaattatggg	901
DB	842	ctgtgtcttggaaggacagctacagttacggaatttgaaccttcagaagaactggg	901
QY	902	ctgggaataacacttcagcttggtgggaataacatgtaattggtcatatgtttaag	961
DB	902	cggtgcaacatttaactgtattgtgggaatagatgtgccagggcgatgcttaacgg	961
QY	962	aaacttcgagattaagaccggtttttacaagt-----gsgagatctgattggagaactt	1015
DB	962	aaacctccaacacatccacggtatttacaagatgataagatggttacccttgaaccgcca	1021
QY	1016	ctccgaacacttttatgatttgaagtagatgcaaggaatggaatgagcttga	1075
DB	1022	ctgtactactccttgatgacttcttggaagaagacacaaagggtgcatggaagcaacgta	1081
QY	1076	ggcgaggtccttggtgctgacagatgtgtcttcaagggaatcatcttttaaaagcgg	1135
DB	1082	ccggtgacctgtggtccgtaacagatgggtgtctcaggttaacctcttaaaagccg	1135
QY	1136	ataccaattgtcgtcaaatatcaacattacttacaatgatttcaaaaatagctttac	1195
DB	1142	atgtaatttaacctcaaatcttccacatgaaactgtatagttgaacaaagtcgtaactac	1201
QY	1196	agcagaatgtagtgaatacagaactactgttaacaatggaatgggaagtgttgaatg	1255
DB	1202	acggatactccacgtccaacgcaacccgttcgcgaatgcaattggggaaatcccgtaatg	1261
QY	1256	gaatgactgcgcttcttattgtgaagaacaggaatgttatcatagtgatctctgaagct	1325
DB	1262	gcattgattggaagactttagtgaagaacggctctgattatatacatgcaatgcttct	1325
QY	1316	ttgtgcgattaaaggttct-----atctggaacaagctatcttagttatataacatgct	1372
DB	1322	ttacacgtatcaatatacctcaactcgaacacagcttatactttagtttaacaacatcgt	1381
QY	1373	aaatcttctcaagcaaggaa	1395
DB	1382	aaaatatcatcatgttaaggaa	1404

RESULT 7
 Z45342
 ID Z45342 strand: DNA; 1107 BP.
 XX


```

DB      946 ggtccgaatggtatctgtcaaaccttaagaagcagtggttt 990
          ||||| ||||| -|| | ||||| || | ||||| || |||
RESULT   9
245345
245345 standard; DNA; 960 BP.
XX
AC      245345;
XX
DT      27-MAR-2000 (first entry)
XX
DE      DNA encoding a Bacillus sp. mannanase enzyme.
XX
KM      Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KM      endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KM      mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KM      synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KM      plant material degradation; recycled waste paper; paper making pulp;
KM      guar; locust bean gum; mannan containing food; coffee extract;
KM      cleaning composition; machine washing; hard-surface cleaner;
KM      dishwashing; oral; dental; contact lens; body-care composition;
KM      fabric softener; oil well drilling; subterranean formation fracture; ss.
XX
OS      Bacillus sp.
XX
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      1..960
FT      /tag= a
FT      /product= "mannanase"
FT      /note= "not termination codon given"
FT      1..84
FT      sig=peptide
FT      /tag= b
XX
PD      16-DEC-1999.
XX
PF      10-JUN-1999; 99WO-DK00314.
XX
PR      10-JUN-1998; 98US-0111256.
PR      20-OCT-1998; 98DK-0001340.
PR      20-OCT-1998; 98DK-0001341.
PR      28-OCT-1998; 98US-0105970.
PR      28-OCT-1998; 98US-0106054.
PR      23-DEC-1998; 98DK-0001725.
PR      05-MAR-1999; 99DK-0000306.
PR      05-MAR-1999; 99DK-0000307.
PR      05-MAR-1999; 99DK-0000308.
PR      05-MAR-1999; 99DK-0000309.
PR      09-MAR-1999; 99US-0123543.
PR      10-MAR-1999; 99US-0123623.
PR      10-MAR-1999; 99US-0123641.
PR      11-MAR-1999; 99US-0123642.
XX
PA      (NOVO ) NOVO-NORDISK AS.
XX
PI      Kauppien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX      WPI: 2000-105891/09.
XX      P-PSDB; Y54132.
XX
PT      New mannanases for treatment of textiles, plant material and coffee
PT      extract, and in cleaning compositions
XX
PS      Disclosure: Page 227-228; 242pp; English.
XX
CC      The present sequence encodes a mannanase enzyme (also known as
CC      mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC      The mannanase hydrolyses galactomannans. Specifically, mannanases
CC      hydrolyse 1,4-beta-D-mannosidic linkages in mannan, galactomannans,
CC      glucomannans, and galactoglucomannans. The mannanase protein, or
CC      preparations containing it, are used to improve properties of cellulosic

```

```

CC      or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC      sizes or printing pastes). They are also used to degrade or modify
CC      plant materials (particularly recycled waste paper, paper making pulps,
CC      or material containing guar or locust bean gums (thickeners), or to
CC      reduce viscosity of mannan-containing foods or feeds). The mannanases
CC      are also used to process coffee extracts (to inhibit gel formation); in
CC      cleaning compositions (for machine washing of fabrics, as hard-surface
CC      cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC      lens or body-care compositions) where they remove mannan-containing
CC      soils and prevent binding of some soils to cellulotics; and in fabric
CC      softeners. They can also be used in oil well drilling to fracture
CC      subterranean formations.
XX
SQ      Sequence 960 BP; 257 A; 183 C; 280 G; 240 T; 0 other;
Query Match      21.1%; Score 310.8; DB 21; Length 960;
Best Local Similarity 59.3%; Pred. NO. 2,5e-77;
Matches 547; Conservative 0; Mismatches 372; Indels 3; Gaps 1;
DB      45 actatagtagctctatctctgttcgttcaggaaaccttacaagtaatgcaattccgg 104
DB      33 actagtggtggtgtgtgtttgtttgtatcacgtagcggtttagatcctgcacaaagcg 92
DB      105 atttatgtaagcgtgtaccactctatagatgcgaatggaaccatttgtatgaggg 164
DB      93 ctltacgtaaaaggatcacgaggtgtgtgacaanaatgvcgactcttcgttatcgtgtg 152
DB      165 gattaacatggcgcgcgcgtatgataaagaccagcaactacatgcaatgaaaggtatgc 224
DB      153 cgtcaaccatgtacacttctgttttaacaagaatttagaggggaatccctccatagc 212
DB      225 aataccggtgtctaaacggtcgcgattgttctatctatggtggggaacatgagcaaaaga 284
DB      213 agaaacaggggcgaacacagatgagatcgtcttaccatgacagcagatggaagaaaga 272
DB      288 tgaatcatcacgtaagaagaacctatctctttagcgaagataatcttggttgcgt 344
DB      273 tgaatcctctgagcttgcgcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 332
DB      345 tcttgaattcatgtaacccgtgtatgatactcattcttcgcacatcgtgcgtgtga 404
DB      333 gctggaagtcaccgactgtctacaggaagtgatatccgatgttagataaagcgtgcga 352
DB      405 ttattgattgaatgagaaagtgcttaatttggaaaggaagataccgctcatatataat 464
DB      393 ttactgtagatgaaatgctgtagtcttcaagaaggagcagacccggtatcatcaat 452
DB      465 tgcgaatgaatggttgcgttgcgtggaagggatgcttgcgtgcggtgtataaacacac 524
DB      453 tgcgaatgaatggttgcgttgcgtggaagggatgacgttgcgtggaagggatgacacac 512
DB      525 aatcccgcatgtgcgtaacgcgcgtgtcaaacatcccttgatgtagatgcgtgcgggtg 584
DB      513 gatcccgcatgtgcgtaacgcgcgtgtgcgtcccatcgttaaatgattgtagcgcggtg 572
DB      585 gggacaatttcacacatcgatcatgatatgaaagaaagttttaatgctgcacctca 644
DB      573 gggacagtagtgcctgtctatccatgagcgggagcagcagattgttcctccgacatt 632
DB      645 acgaataacatggttttcgatcatatgtatgatatgacaggtgtatgataatgcagatc 704
DB      633 aaaaacacaaatggttttcgatcatatgatatgacaggtgtatgataatgcagatc 692
DB      705 tctactaatatgtgacacgattcttaacaaagcctcgaattgcatgcatggtgaaattgg 764
DB      693 ttctgaaaaacgtcgaacggtgtactctgtgaaatcttctgtgtaacgcgtgaaattgg 752
DB      765 acacgcgtatataaattgtgacgttcagatgaaagcaagataatgagctattctgaaacaag 824
DB      753 catagcactcatgtagcgtatgcagatgaaagatcgattttgacctatacagcagagcg 812
DB      825 aggaattggtgtgtgtgcgtgtcatgtaagaagggagccagaaatgtagattataga 884

```

[illegible]

CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to cellulose; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.

CC Sequence 397 BP; 113 A; 69 C; 116 G; 99 T; 0 other:

Query Match 7.1%; Score 104.4; DB 21; Length 397;
Best Local Similarity 57.3%; Pred. No. 8.3e-20;
Matches 189; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

OY 94 gaaattccggtatttattgaagcggtaccactctataagatgccaatggaaccattt 153
DB 67 ggcacaactggatttccaagtaacaggagaccacggttgcttgatggagggcaatccgtat 126
OY 154 gtaatgagagagattaacattgagcagcatgtgataaagaccaggaactactgcaatt 213
DB 127 gtgatgctgtggatcattacatcagcagcatattgttcaaaacaggaacttgatacagaata 186
OY 214 gaaggatgccaataaccggtgataatacagtcgcgattgtgtatctgattggggagaca 273
DB 187 ccagtatgtcagcagcatcgtcgtatatacagtgagagatcgtttatcgtatggccaaca 246
OY 274 tggacaagaatgagatcatcatcactaagaacctatctcttggcggaaataatctt 333
DB 247 tggagcgagataaccgttgcggagattgaagaagtcttcagttaccgaagaggaagcg 306
OY 334 ttggtctgcttcttgaagtcattcatgctcaggttatgattcattcttcgcctcaat 393
DB 307 ttgaagcgctgattcttgaagttcattatgctgacggaggaagtgtatgaccaaacgattgttt 366
OY 394 cgtgctgtgattatttgattgaatgaga 423
DB 367 actgcagtgagattgtgtcagagagagaga 396

RESULT 14
T83966
ID T83966 standard; DNA: 2635 BP.

AC T83966;
XX
DT 26-AUG-1998 (first entry)

XX DNA encoding a mtla gene enzyme and a protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW Staphylococcal gene; regulatory element; bacterial gene expression;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome; ss.

XX Staphylococcus aureus.

OS
XX
FH Key Location/Qualifiers
FT CDS 339..1004
FT /*tag= a
FT /product= mtla gene enzyme
FT 1193..1750
FT /*tag= b

XX W09730070-A1.

XX 21-AUG-1997.

PD 19-FEB-1997; 97WO-US02318.

XX

PR 20-FEB-1996; 96US-0011888.

XX (SMIR) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI; 1997-424969/39.

DR P-PSDB; W28008, W28009.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used

PT to isolate antimicrobial compounds, and in vaccines against S.

PT aureus infection

PS Claim 9; Page 778-779; 989pp; English.

CC The present sequence encodes a Staphylococcus aureus protein, that
CC based on homology with a S. carnosus protein is believed
CC to be a mtla gene for enzyme II (Mannitol), and another protein
CC that is of unknown function. The present sequence was obtained from a
CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC sequence can be used in the construction of ribozymes and antisense
CC sequences to control the expression of Staphylococcal genes. The DNA
CC sequence is also useful as a source of regulatory elements for the
CC control of bacterial gene expression. The encoded protein may be used
CC to produce vaccines to enable a host to produce specific antibodies
CC with antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.

XX Sequence 2635 BP; 886 A; 437 C; 474 G; 834 T; 4 other:

Query Match 2.8%; Score 41.8; DB 18; Length 2635;
Best Local Similarity 55.0%; Pred. No. 0.059; Mismatches 67; Indels 0; Gaps 0;
Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 886 attggcctggaataacccttaccagcttgggaataacatagtaattgtccatattg 955
DB 376 atggtcgtggaattatccactcttagtggatcattcgaattattcccatattgtat 435

OY 956 taagagaacctcgagattaagacgcgttttaacaggtggagagatcattgaggaactt 1017
DB 436 taatgcgtccttattattatcgtctgtatttagtgaggaatgactgtgtgaactt 405

OY 1016 ctcgcgaacctttatgatttgaagt 1044

DB 496 atcaagcaacaggaatttgattttaaagt 524

RESULT 15
V74586/c
ID V74586 standard; DNA: 4826 BP.

AC V74586;
XX
DT 16-MAR-1999 (first entry)

XX Staphylococcus aureus contig SEQ ID #275.

XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.

XX Staphylococcus aureus.

OS
XX
FH Key Location/Qualifiers
FT misc_feature 1081..1140
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They

Search completed: December 20, 2000, 09:58:32
 Job time: 30586 sec

```

FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence
FT      misc-feature      2881..2940
FT      /tag= b
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
FT      misc-feature      4681..4740
FT      /tag= c
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
XX      EP786519-AZ.
XX      PD      30-JUL-1997.
XX      PF      07-JAN-1997; 97EP-0100117.
XX      PR      05-JAN-1996; 96US-0009861.
XX      PA      (HUMA-) HUMAN GENOME SCI INC.
XX      PI      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX      PI      Rosen CA;
XX      DR      WPI; 1997-374922/35.
XX      PT      Polynucleotide(s) and proteins derived from Staphylococcus aureus
XX      PT      stored on computer readable medium and used in the production of
XX      PT      anti-S.aureus vaccines
XX      PS      Claim 1: Page 1089-1092; 3271pp; English.
XX      CC      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX      CC      of the invention. The DNA sequences are recorded on a computer readable
XX      CC      medium, preferably selected from a floppy or hard disk, random access
XX      CC      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX      CC      the S.aureus DNA sequences allows putative functions to be assigned so
XX      CC      that protein encoding or regulatory regions of commercial, therapeutic or
XX      CC      industrial importance can be obtained. Specifically, sequences which are
XX      CC      likely to encode antigens have been identified and these polypeptides can
XX      CC      be used in a vaccine composition against S.aureus infection. The
XX      CC      polypeptides can also be used in a kit for the immunodetection of
XX      CC      S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX      CC      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX      CC      skin and surgical wound infections, scalded skin syndrome, toxic shock
XX      CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
XX      CC      for recombinant production of the polypeptides. The new DNA sequences
XX      CC      (and their fragments) are useful as primers or probes for isolating
XX      CC      homologues of any of the S.aureus DNA sequences contained on the
XX      CC      computer readable medium.
XX      SQ      Sequence 4826 BP: 1645 A; 707 C; 893 G; 1400 T; 181 other;

Query Match      2.8%; Score 41.8; DB 18; Length 4826;
Best Local Similarity 55.0%; Pred. No. 0.077;
Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY      896 attggcgtggaataacttaacgtcttgagggaataacatagtgatagtcatatgtt 955
OY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      348 ATGGTGCTGGAAATTCACCTCTTAGGTGATTCATGAATTTATTTCCATATGAT 289

OY      956 taagagaactcgaagactaagcacgltttacagtgaggagatcgtatggagaactt 1015
OY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      288 TAATGGCTCTTATATATTATCGCTGATTTAGTGATGACTGCTGTGCAACTT 229

OY      1016 ctccgacaactcttatgatttgaagt 1044
OY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      228 ATCAAGCAACAGGATTTGATTTAAAGT 200
  
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 01:23:33 ; Search time 75.74 Seconds
(without alignments)
2935.361 Million cell updates/sec

Title: US-09-339-159-1
1470

Perfect score: 1 ttgataatggttttaaaa.....ttgataatggttttagaa 1470

Sequence: 1 ttgataatggttttaaaa.....ttgataatggttttagaa 1470

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/1/lna/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/lna/5C_COMB.seq: *
4: /cgn2_6/ptodata/1/lna/5D_COMB.seq: *
5: /cgn2_6/ptodata/1/lna/6_COMB.seq: *
6: /cgn2_6/ptodata/1/lna/PCRTUS_COMB.seq: *
7: /cgn2_6/ptodata/1/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48.6	3.3	7218	1	US-08-232-463-14
C 2	33.6	2.3	51259	5	US-08-781-891-209
C 3	33.4	2.3	1365	4	US-08-870-827-4
C 4	33.4	2.3	2408	4	US-08-870-827-5
C 5	33.2	2.3	3138	1	US-07-867-106-4
C 6	33.2	2.2	782	1	US-08-261-825-1
C 7	33.2	2.2	782	1	US-08-719-124-1
C 8	33.2	2.2	782	6	PCR-US95-07748-1
C 9	33.2	2.2	782	6	PCR-US95-07748A-1
C 10	33.2	2.2	1800	5	US-09-039-773A-1
C 11	32.4	2.2	387	1	US-08-222-177A-14
C 12	32.4	2.2	5467	1	US-07-745-206A-12
C 13	32.4	2.2	5467	3	US-08-311-363-12
C 14	32.2	2.2	3030	4	US-08-680-326-24
C 15	32.2	2.2	8133	2	US-08-480-604A-5
C 16	32.2	2.2	8133	4	US-08-405-496A-5
C 17	32.2	2.2	4339	4	US-08-566-398-23
C 18	32.2	2.2	4365	6	PCR-US91-08525-21
C 19	32.2	2.2	4365	6	PCR-US91-08525-25
C 20	32.2	2.2	4365	6	PCR-US93-04384-1
C 21	32.2	2.2	4365	6	PCR-US93-04384-7
C 22	32.2	2.2	4365	6	PCR-US93-04384-15
C 23	32.2	2.2	6090	3	US-08-566-398-35
C 24	32.2	2.2	6144	3	US-08-566-398-32
C 25	31.8	2.2	7366	7	5169760-3
C 26	31.8	2.2	7852	5	US-08-836-022A-2

27	31.8	2.2	7897	5	US-08-836-022A-1	Sequence 1, Appl
C 28	31.8	2.2	8509	2	US-08-462-014-1	Sequence 1, Appl
C 29	31.8	2.2	10398	3	US-08-331-384-1	Sequence 1, Appl
C 30	31.8	2.2	10398	3	US-08-708-188-1	Sequence 1, Appl
C 31	31.8	2.2	10398	3	US-08-836-087-1	Sequence 1, Appl
C 32	31.8	2.2	19182	4	US-08-850-880-11	Sequence 11, Appl
C 33	31.8	2.2	19182	4	US-08-944-916-11	Sequence 11, Appl
C 34	31.6	2.1	40328	5	US-08-742-185-102	Sequence 102, App
C 35	31.6	2.1	43795	5	US-08-742-185-101	Sequence 101, App
C 36	31.4	2.1	8855	3	US-08-542-003-1	Sequence 1, Appl
C 37	31.4	2.1	8855	3	US-08-322-760A-1	Sequence 1, Appl
C 38	31.4	2.1	1453	4	US-08-834-108-1	Sequence 1, Appl
C 39	31.4	2.1	1453	4	US-08-252-995D-1	Sequence 1, Appl
C 40	31.4	2.1	1600	1	US-08-834-108-5	Sequence 5, Appl
C 41	31.4	2.1	1600	4	US-08-834-108-5	Sequence 5, Appl
C 42	31.4	2.1	2882	1	US-08-424-567-1	Sequence 1, Appl
C 43	31.4	2.1	2882	3	US-08-711-928-1	Sequence 1, Appl
C 44	31.4	2.1	3447	1	US-08-252-995D-3	Sequence 3, Appl
C 45	31.4	2.1	3447	4	US-08-834-108-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFFELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELE: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
; US-08-232-463-14

Query Match 3.3%; Score 48.6; DB 1; Length 7218;
 Best Local Similarity 3.8%; Pred. No. 0.00022;
 Matches 15; Conservative 216; Mismatches 160; Indels 0; Gaps 0;

```

QY 747 agcatgtgtgaattgacacgcgtcatcaaatgtgacgtgacgaacgattat 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1455 AGGATAGAGAAAGATTGTGACRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1396
QY gaactatctgcaaaagagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 866
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1395 RRRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1336
QY 867 agaatgagagatattagacattcgaaatgattgctgtgaaataacattacgtggg 926
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1335 RRRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1276
QY 927 aatataagtagtgatgcatatgttttaagaagaactcgaaatlaagcagctttt 986
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1275 RRRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1216
QY 987 tacaggctgagagctctgatgagaaactctccgacaactcttaattgaaagtag 1046
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1215 RRRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1156
QY 1047 tatgcaagagatgactggaagtagcttgagcggagtgctcttggtgtgacagatggtc 1106
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1155 RRRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1096
QY 1107 ttctaagaagatcattctttaaagcgat 1137
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1095 RRRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1065
  
```

RESULT 2

```

US-08-781-891-209/c
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51259 base pairs
  
```

TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-781-891-209

Query Match 2.3%; Score 33.6; DB 5; Length 51259;
 Best Local Similarity 59.4%; Pred. No. 11;
 Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```

QY 3 gataatgttttaaaaaatttttctataacattactactactagctactat 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 29076 GAAAAAGCTGCTTAATAATTATTCTTAATCATAGCTATCTGCTAAGATTC 29017
QY 63 tctgttgcttcaagaaactctacagctaatgc 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 29016 ATTGTTTATTATTACTGTTACAGATGAGGAAA 28981
  
```

RESULT 3

```

US-08-870-827-4
; Sequence 4, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Terasaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1365 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-870-827-4
  
```

Query Match 2.3%; Score 33.4; DB 4; Length 1365;
 Best Local Similarity 55.7%; Pred. No. 2.6;
 Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```

QY 262 gatggggacaatgacaaagatgacatccatagtaagaacctatctcttagcg 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 655 GATGAAGCAAGTGAATTCAGGTGATTAACCACTTTAAGAGACCCCTACATGTGAA 714
QY 322 gaagataatcattgtgtcttcttgaaatcatgatgctacggttatgatt 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```


GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,825
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-261-825-1
Query Match
Best Local Similarity 57.1%; Score 33; DB 1; Length 782;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 1365 aaatgccaaaatcttccaaagtaaggaattggagtcagttccaatcagcgagtaga 1424
DB 35 AAATTACCAAAATCTTAATCAACAATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1425 tagtagtgacaaacatcgattatattgataaagtgttaga 1469
DB 95 TAGCTCTGTCTAATCTTAATGATTGCTAATAGAAATGATA 139
RESULT 7
US-08-719-124-1
Sequence 1, Application US/08719124
Patent No. 5854395
GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,825
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-719-124-1

CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,124
FILING DATE: 24-SEP-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,825
FILING DATE: 17-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-719-124-1
Query Match
Best Local Similarity 57.1%; Score 33; DB 3; Length 782;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 1365 aaatgccaaaatcttccaaagtaaggaattggagtcagttccaatcagcgagtaga 1424
DB 35 AAATTACCAAAATCTTAATCAACAATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1425 tagtagtgacaaacatcgattatattgataaagtgttaga 1469
DB 95 TAGCTCTGTCTAATCTTAATGATTGCTAATAGAAATGATA 139
RESULT 8
PCT-US95-07748-1
Sequence 1, Application PC/TUS9507748
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

Query Match Similarity      2.2%: Score 33; DB 6; Length 782;
Best Local Similarity      57.1%: Pred. No. 2.6;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0.

QY 1365 aaatlyccaanaattcttctcaagtaagggaattgagttcagttccaatcagcgagtga 1424
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35 AAATACCAAAATCTTAATCAACATGCTTAATTGGCAACCAATAAGACTATTGT 94
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1425 tagtagtggacaacaacatcgatttatatgatataatgtatgtatga 1469
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 TAGTCTGTCTAAATCTTAATTAGTATGCTAATAAGATGTATA 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
PCT-US95-07748A-1
; Sequence 1, Application PC/TUS9507748A
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: CLONED BORRELIA burgdorferi VIRULENCE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07748A
; FILING DATE: 16-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HAITE, Ph.D., LISA A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07419/013WO1 (PD3516)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

Query Match	2.2%	Score 33	DB 6	Length 782
Best Local Similarity	57.18	Pred. No. 2.6		
Matches 60	Conservative 0	Mismatches 45	Indels 0	Gaps 0
QY 1365	aatgctcaaatcttcctcaagtaaggaatctggagttcagttccaatcagcgagtga	1424		
DB 35	AAATTACCAAAATTTTAAATCAACATGCTTAATTTGCCAACCCAAATTAAGCTATTTTGT	94		
QY 1425	tagtagtcggacaacatcgatattatattgacaatcgtgatttaga	1469		
DB 95	TAGCTTGTTCTTAATTTCTATTAGATATGCTTAATTAAGATGATAT	139		

RESULT 10
US-09-039-773A-1
Sequence 1, Application US/09039773A
Patent No. 6100388
GENERAL INFORMATION:
APPLICANT: Casas, Ivan
APPLICANT: Jonsson, Hans
APPLICANT: Mvilstam, Bo
APPLICANT: Roos, Stefan
TITLE OF INVENTION: Lactobacilli Harboring Aggregation and Mucin
TITLE OF INVENTION: Binding Genes As Vaccine Delivery Vehicles
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Standley & Gilcrest
STREET: 495 Metro Place South, Suite 210
City: Dublin
STATE: Ohio
COUNTRY: US
ZIP: 43017
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS Version 6.22
SOFTWARE: Microsoft Word Version 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,773A
FILING DATE: 16-Mar-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA: No. 6100388 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Donald O. Nickey
REGISTRATION NUMBER: 29,092
REFERENCE/DOCKET NUMBER: 1229-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 792-5555
TELEFAX: (614) 792-5536
TELEX: No. 6100388 applicable
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Circular
MOLECULE TYPE: Genomic DNA
DESCRIPTION: Genomic DNA sequence and deduced amino
DESCRIPTION: acid sequence of bacterial aggregation
DESCRIPTION: protein

```

; MOLECULE TYPE: DNA (genomic)
;
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Caucasian
;

```

Oy	17	aaaaatttttcacatcaatcaatcaatgtagctcattcgttttcag	/b
Db	113	AAAACATTTTAAAACTTGAATTCGATTCGCAAAATCCCTATCCTTGTCTG	172
Oy	77	gaacttcacagcgcaatcgcaatccggaattttatgcaagcgytacaactctaatcgtg	136
Db	173	TAAAGACTAAACAAATTCACAGAGCACTGTTAAACAAACAAATGAAGTGAACCT	232
Oy	137	ccaatgaaaccacattgtaatsga	162
Db	233	ACCATTGAAGAAAGAACATTAAGGATA	258

APPLICANT: ROSE, TIMOTHY M.
APPLICANT: BOSCH, MARINX
APPLICANT: STRAND, KURT
APPLICANT: TODARO, GEORGE J.
TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
TITLE OF INVENTION: FIBROMATOSIS
NUMBER OF SEQUENCES: 152
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FORRESTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,326
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Schif, J. Michael
REGISTRATION NUMBER: 40,253
REFERENCE/DOCKET NUMBER: 29938-20001.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 3030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-680-326-24

Query Match 2.2%; Score 32.2; DB 4; Length 3030;
Best Local Similarity 53.6%; Pred. No. 8;
Matches 67; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 51 agctagctatcttctgttcttcaagaaacttcaagcgaattccggaattta 110
DB 963 AACTAGAAATATGCTATTATCTTAGAGACGTGCTGAGAAATCTGAACTTAA 1022
QY 111 tgaagaggtaaccactctatagcgtgcgaatggaaccacttctgaagaggaattaa 170
DB 1023 TGAGTTCCTTAGGAAATAGACATGTCATGGGTTTTTTCATTATATAGAGACTTAA 1082
QY 171 ccactg 175
DB 1083 TGTGTG 1087

RESULT 15
US-08-480-604A-5
Sequence 5, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
NUMBER OF SEQUENCES: 32
PREVENTION OF C. DIFFICILE DISEASE
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8133 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..8130
US-08-480-604A-5

Query Match 2.2%; Score 32.2; DB 2; Length 8133;
Best Local Similarity 49.1%; Pred. No. 12;
Matches 85; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 2 tgataatggttttaaaaaatttttctataacattatcattactatagctagctcta 61
DB 5333 TGAGTATAGATTTTAAAGATATTAATAAACTATCATTTAGATATTAATGAATTTTA 5392
QY 62 ttctgttcgtttcaagaaacttcaagcgaattccggaatttctgaagcggtta 121
DB 5393 AATCATTTAATCTGAAATATGAAATTAAGATTAAGATCATTTAGCATTTAAATTAATGATA 5452
QY 122 ccactctatagcgtgcgaatggaaccacttctgaagaggaattaaaccat 174
DB 5453 ATAAACTATTACTATGATGAGATGATGAATTAATTAAGTTAAAGATTAATCAAT 5505

Search completed: December 20, 2000, 09:56:10
Job time: 30757 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 19, 2000, 23:01:31 ; Search time 1967.85 Seconds
(without alignments)
4618.607 Million cell updates/sec

Title: US-09-339-159-1
Perfect score: 1470
Sequence: 1 ttgaataatggtttaaaaa.....ttgaataatgtattgtagaa 1470

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

DB	954	KTAGAGCGDKKKKKDKNDKRDADATGDKDWTWBDTDTOTAKRRRRRRRRRAGGGRDKD	1013
Oy	1062	tygaagtgcttgagcggaggtccttggtcgttgacagagtgcttcctcaaggagta	1121
Db	1014	KKRRRAARRRRKKKKTKDKDDGGRWRARAKADAAKADADAADAAAMAKAKAKDK	1073
Oy	1122	ttcttcaaacgagat	1139
Db	1074	DKAKKADADDKDKKAK	1091
RESULT	2		
CNS00396/C			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match	3.1%	Score 46.2;	DB 121; Length 1101;
Best Local Similarity	14.5%	Pred. NO. 0.069;	
Matches	65;	Conservative 215;	Mismatches 164; Indels 3; Gaps 1;
Oy	773	atacaaatgtagctgtagatgaagaacagattatgacatctctgacaaagagagt	832
Db	1026	WWWWATAKTDTAAWTWWRKRAMADNAGNDRGGRKRDADATADAGGRDGGRRKRDND	967
Oy	833	ggtggttgcgtygtcatgaaaggaagcagcccaagatggagattatgaccttga	892
Db	966	RKDGDGDKKGGKKRAAKAAKAWATKMWDDMDMDKMKWGDADRRKDDDDGADKDDDK	907
Oy	893	atgattggctgtagaataaccttaacgctctgggggaataataatgtagtgcata	952
Db	906	GADADDGDTGTDKDDDKDKMDKAKGTGATWAAATDWWGWAADADWWTDAAD	847

```

QY 953 gtttaaggaaactcgc---gattaagcacccgttttaacagggtgagagatcgtltag 1009
DB 846 DMMADRRDMDAMWKKDDAMWAGFARADRRDMDRAGKRGAGKRRDRADRRADRAADR 787
QY 1010 gaactctccgaacactcttatagttttgaagtagtataagaaagtcgaagta 1069
DB 786 DDAATWTTWTTTTTTDDTDDMKKKTCTTWRKMAADRWDADDDDDRRAGTAGKWRRTK 727
QY 1070 gctttagcaggaggctccttggtgagcgtgacagagtgcttctaagaagtcattc 1129
DB 726 RWRKRRDTRWDADADTADTADRRRRRGDDGADAGKCKTGKRRRRDRATWPTDAMADA 667
QY 1130 aagcgatattcaatcgtcgtcgaataactcacacattactatcatgttttcaaaatag 1189
DB 666 AAWTTTDDTDDTDDMDRRRRKRRRRRTTAAADMDWTWAMDAKKMDKTRADRMWR 607
QY 1190 cttaacagcagatagtagatacaag 1216
DB 606 AADTWTDAKRRADRDWAKARAWARRDR 580

RESULT 3
CNS00LT2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
BASE COUNT
ORIGIN

CNS00LT2 1101 bp DNA GSS 14-JUN-1999
Drosophila melanogaster genome survey sequence TE33 end of BAC:
BACR48B19 of RPci-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL078714
AL078714.1 GI:5102004
GSS.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 51006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPci-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
15ogenic strain Y2; cn bw sp, the same strain used for the Blue
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1lb="RPci-98"
/clone="BACR48B19"
/note="end : TE33"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

ORIGIN
Query Match 3.0%; Score 43.4; DB 121; Length 1101;
Best Local Similarity 14.9%; Pred. No. 0.4;
Matches 48; Conservative 140; Mismatches 135; Indels 0; Gaps 0;

315 tttagcgaaataatcatcttggtcgtcttcttggaatcaatgactaccggtatga 374

```

```

Db 699 TTKKKKKKAAADKKWDKAKKATKTKKDKKAAADKKDRKKGGKGGKGG 758
      ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Qy 375 ttcattcttcgtcgaatcgtctgttattgattgaaagagtgcttatt 434
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 759 KKGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 818
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 435 tgaagaagaataccgcatattatattgaaagagtgcttattgaaag 494
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 819 KKKADRTKTKTKYDAAAAAAAKTKDKKKKTKTKTKTKTKTKTKTKTK 878
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 495 ggaatcctggagtcgagtgatataaagaagcaatcccgatctgtaagcgcgtctaa 554
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 879 KKKKGTKKKKKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 938
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 555 ccaatccttgatgagtagtctgcggtggtgggacaattccacaatgattcatgatta 614
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 999 KKGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 998
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 615 tgaagaagaagtttcaatgctg 637
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 999 KTKKKDAAAAATKKKKKKKKKKTK 1021
      : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
CNS0182P 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108811.1 GI:5629115
VERSION AL108811.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
          Genoscope.
          Direct Submission
          Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the European Drosophila Genome Project (EDGP) -
          http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
          library (Dros BAC) was made by Alain Billand at CERN (Centre
          d'Etude du Polymorphisme Humain) with funding provided by a MRC
          project grant. The DNA was prepared from embryos by Alain Bucheton
          and Genevieve Payan. It has been constructed in the vector
          pBelobAC11.
FEATURES
  source
    1..1101
    /organism="Drosophila melanogaster"
    /plasmid="pBelobAC11"
    /db_xref="taxon:7227"
    /clone_1lb="DrosBAC"
    /clone="BACN37D10"
    /note="end : SP6"
BASE COUNT 274 a 268 c 128 g 73 t 358 others
ORIGIN

```

```

Query Match 3.0%; Score 43.4; DB 121; Length 1101;
Best Local Similarity 16.1%; Pred. No. 0.4;
Matches 66; Conservative 174; Mismatches 169; Indels 0; Gaps 0;
Qy 668 atagtataatgagtgatgtaagtcgacgaagtcgtactataatgaccgaagtc 727
      |:|:| | | | | | | | | | | | | | | | | | | | | | | |
Db 692 ADKRRRAAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 751
      |:|:| | | | | | | | | | | | | | | | | | | | | | | |
Qy 728 ttaataagacccgcatgattgattgaaatggaacgcgtacataatggtgagc 787
      |:|:| | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 752 WWTWDATWADTWAAADTDAKRAAAGRRKRDARTARRRARRRAAAGGKRRABAG 811
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 788 tcatgaagaacagattatgagctatcgaacaagaagtggtggtggtggtggt 847
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 812 ARARARARARARARARARARARARARARARARARARARARARARAR 871
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 848 catgaaagggaaagcccaagatggaattgagaccttcgaatgattggtcgaga 907
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 872 DARARARRRRRRRRRRARRARRARRAADTDTKDRNADATDTKDTTDTDDDMKAK 931
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 908 atacccttcagcttggtggaataacacatgtaagtgctcatatggttaagaagaact 967
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 932 RDRWMAAKADGAWKWRDRARADMAATKDDGKRWGGRGGRGKDKKRWKDKTKGKD 991
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 968 cgaattaagcacccgttttacagtgtagagatctgattgagaacttcgcgaactc 1027
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 992 DDDWDKTTWRDMMWMTTKKMDMMWDGGRGRTWRKRGGAWWRADWARRDRTDGRDTRT 1051
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1028 ttaatgatttgaagtagtatgcaagtagtgaagtagtggag 1076
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 1052 ADKRRRTDTTKRDGDWRRKDRKDKGDKDKDKDKDKDKDKDKDKDKDKDKDKDKDK 1100
      : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 5
CNS00ES1/c 1101 bp DNA GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TERT end of BAC:
DEFINITION BACR29L15 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL069797.1 GI:4949738
VERSION AL069797.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
          Genoscope.
          Direct Submission
          Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Osoegawa and
          Aaron Mammosser in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCT-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain y2; cn bw sp, the same strain used for the BACPAC
          PI and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
  source
    1..1101
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_1lb="RPCT-98"
    /clone="BACR29L15"
    /note="end : TERT"
BASE COUNT 174 a 179 c 72 g 207 t 469 others
ORIGIN

```

```

Query Match 2.9%; Score 42.8; DB 121; Length 1101;
Best Local Similarity 19.7%; Pred. No. 0.58;
Matches 122; Conservative 196; Mismatches 300; Indels 0; Gaps 0;

```


ORIGIN

Query Match 2.7%; Score 40.2; DB 15; Length 631;
Best Local Similarity 47.8%; Pred. No. 2.5; Mismatches 128; Indels 0; Gaps 0;
Matches 117; Conservative 0;

893 atgattggcgtgaaataacttaacagcttggggaataacataagtgatggtccatag 952
|||||
430 atgattgacttgcgttctgtgtgacagcttgggttgcacacatttgatgacgttg 371
|||||
953 gtttaagaagaacttcgagattgaacacgcttttaacagtgtagatctgtagaaga 1012
|||||
370 gttgttcacagacattcagctgacacttctgtgcttgcctgacactgattcagctg 311
|||||
1013 ctctccgacacactttatgatttttgaaagtagatgcaagtagagtagt 1072
|||||
310 gttgttgccttgcacagctgacagctgacagctgacagctgacagctgacagctg 251
|||||
1073 tgaagcgaagctccttggcgtgacagagtggtctcttaagaagtagctcttaaga 1132
|||||
250 cggcagacgtatgattgacgtgacagctgacagctgacagctgacagctgacagctg 191
|||||
1133 cggat 1137
|||||
190 ctgat 186

RESULT 12

LOCUS A032958 552 bp DNA GSS 06-MAR-1999
DEFINITION HS_5001_A2_P09_T7 RPII1 Human Male BAC Library Homo sapiens
genomic clone Plate=577 Col=18 Row=K, DNA sequence.

ACCESSION A032958
VERSION A032958.1 GI:4130585
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 552)
Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 577 row: K column: 18
Seq primer: T7
Class: BAC ends
High quality sequence stop: 552.

FEATURES

source
1..552
Location/Qualifiers

BASE COUNT 177 a 133 c 67 g 170 t 5 others
ORIGIN
1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=577 Col=18 Row=K"
/clone_lib="RPII1 Human Male BAC Library"
/sex="Male"
/cell_type="Lymphocytes"
/note="vector: pBac3.6; RPII1 Human Male BAC Library"

Query Match 2.7%; Score 39.4; DB 96; Length 552;
Best Local Similarity 53.7%; Pred. No. 4; Mismatches 68; Indels 0; Gaps 0;
Matches 79; Conservative 0;

294 tacagtaagaaccttacccttaagcgaagataacattgtgtgtcttctgaagt 353
|||||
182 taaatgaaatattattattatgacgataagaaatattattcttgcctctgattatgt 123
|||||
354 tcatgactcaccggttattatgattccattgcttcgctcaatcgctgttattattgat 413
|||||
122 ccaagactgcagcgtttgtgtgttgaaatgcattcattcattcattcattcattcatt 63
|||||
414 tgaatgagaagtgcttatttgaaa 440
|||||
62 tgatgagaaatgattaaatcagcaga 36
|||||

RESULT 13

LOCUS A0621653 520 bp DNA GSS 16-JUN-1999
DEFINITION HS_3089_B2_H12_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3089 Col=24 Row=P, DNA sequence.

ACCESSION A0621653
VERSION A0621653.1 GI:5084045
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 520)
Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3089 row: P column: 24
Seq primer: T7
Class: BAC ends
High quality sequence stop: 520.

FEATURES

source
1..520
Location/Qualifiers

BASE COUNT 146 a 100 c 129 g 143 t 2 others
ORIGIN
1..520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3089 Col=24 Row=P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

Query Match 2.7%; Score 39.2; DB 100; Length 520;
Best Local Similarity 57.3%; Pred. No. 4.5; Mismatches 53; Indels 0; Gaps 0;
Matches 71; Conservative 0;

313 tcttttagcgaagataacattgtgtgtcttgaagatcatgactcaccggtat 372
|||||
211 tcttttcgacatgacagacttcttatttttttttttcgttttagcttcacagatttt 270
|||||
373 gattccattgctgcctcaatcgtgctgtgattatttgattgaaatgagaagtcttta 432
|||||

Db	271	GTTTAATGATTGGTTCAGGATTCGGTTCAAGATTGCACATAAAGAAGATGAATTAA	330
Oy	433	attg 436	
Db	331	AATG 334	
<hr/>			
RESULT 14			
CNS016e2/c			
LOCUS			
DEFINITION			
Drosophila melanogaster genome survey sequence T7 end of BAC			
BACN15AI2 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
AL106628			
GI:5622852			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
fruit fly.			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : segreifgenoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (BDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.			
FEATURES			
Source			
Location/Qualifiers			
1..1204			
/organism="Drosophila melanogaster"			
/plasmid="pBelobAC11"			
/db_xref="taxon:7227"			
/clone_11b="DrosBAC"			
/clone="BACN15AI2"			
/note="end : T7"			
BASE COUNT			
ORIGIN			
298 a 172 c 106 g 316 t 312 others			
Query Match 2.7% Score 39.2; DB 121; Length 1204; Best Local Similarity 19.6%; Pred. No. 5.7; Matches 65; Conservative 134; Mismatches 131; Indels 2; Gaps 1;			
Oy	1135	gatcatcaatctgtcgaactcacacacttcttacatggttatccaatatcgcttta	1194
Db	1181	KATAKAKMKKAKAKKAAAAAAAADAAAAAADA	1122
Oy	1195	cagcagaatag--taggatacaagtactgttaaacatgcaaatggggaagtgttgaa	1252
Db	1121	ADDAAAAARAKAKKAAAAAADAAWKKKAAAAAAWMAAKKATKKDDDKDKAA	1062
Oy	1235	atggaatgactgcgcgtctttatgtgaaaacagcagatggtatacatagctctcgaa	1312
Db	1061	DARKKKRKAADADRAADKDCKKAKADADMDADADAKKKKKKKDDATPADWA	1002
Oy	1313	gcttgtgcgcttaaggcttcacctgcagacaagctacctctagattatcaaatgtcc	1372
Db	1001	KAAKKKDKAMKKKKDKKAKKAKKAKKDKDKDKDKKKKKDKKAKKRRRAAKDA	942
Oy	1373	aaaatcttctcaagttaaaggaatgtgaggtccagttccaatcagcgaagtgaagt	1432
Db	941	AAAAAAKKKDKKDAKAAKAKKDKDAKKKKDDMAADAKAAAKKRAAKKKDK	882
Oy	1433	gacaaacatcgatttatattgataatgtgatt	1464

```

Db      981  KMDADKKAKDAAKKKKDAAKKADKDAK 850
RESULT  15
LOCUS   A2138670
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
    source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity 53.6% Pred. No. 5.7
Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0
QY 322 gaagaatacattggttgcgttcttgaagtcacgacggtatgattccatt 381
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
QY 382 gcttcgcccaatcgcgctgttgcattatgagatgaaagagagcttattggaag 441
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 GATTGGCTCAATAGTAAGACGATGATGATGATGATGATGATGATGATGAT 431
QY 442 gaagatccgctcattatgattatgcaatg 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 GAAGATGACGACATGATGATGATGATGATG 462
Search completed: December 20, 2000, 01:58:22
Job time: 10611 sec

```

Wed Dec 20 15:33:30 2000

us-09-339-159-1.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 03:17:03 ; Search time 4256.96 Seconds
(without alignments)
1415.860 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_1470
Perfect score: 1380
Sequence: 1 aatgcacatccgatttta.....ttgataatgtgattgtagaa 1380

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_cm:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_v1:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pl4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum3:*
48: gb_pl3:*
49: gb_pl5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_v11:*
73: gb_v12:*
74: gb_ba3:*
75: em_htg6:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pl6:*
92: gb_pl7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	458.6	33.2	1551	1 AB007123	AB007123 Bacillus
2	242	17.5	1530	2 D86329	D86329 Vibrio sp.
3	134	9.7	4567	1 AF163837	AF163837 Calidpact
4	131.8	9.6	2155	2 CDCMANABD	M3063 C.saccharol
5	131.8	9.5	4977	2 CDCMANA	I01237 Caldocellum
6	125	9.1	5284	2 CSU16308	U16308 Caldocellum
7	109.4	7.9	1461	5 E02075	E02075 Genomic DNA
8	109.4	7.9	1939	2 BACMANN	M31797 Bacillus sp
9	94.4	6.8	25970	2 SC2H4	AL031514 Streptomy
10	89.2	6.5	1881	74 STMBMANASE	M92297 Streptomyce
11	81.6	5.9	837	74 TFG6227	AJ006227 Thermomon
12	52.4	3.8	14922	2 SCF73	AL121746 Streptomy


```

Db      878 TGCGTATCTGATGATGACAAATGATTGGCGTGTACCTCCCTACCTCGTGGTATA 937
Qy      842 caatagtgaatggtccatattggtttaagaagaacctcgagattaagcaccgttttacag 901
Db      938 CCTAGTGAATGACAGTAAAGCACTTCTGTATATCGGCAATTTTGAG 997
Qy      902 gtgagagatctgtgtgaggaactcttcgcgaacatc----- 937
Db      998 GTGTACGCCACCTCAAGCCCTACTTCTACACCTACATCTACGCCCACTCAACTCTTA 1057
Qy      938 -----ttatgatttg 949
Db      1058 CTCCTAGGCCAAGTCGAGCCCGAGTCGAGTAATAAGGAGAGATCTATATGATTTCG 1117
Qy      950 aagtgatgataaggaatgagctggaagtagcttgagcggaggtccttggtgtgacag 1009
Db      1118 AAACAGGAACTCAAGCGTGTGCGGAACAATATTTTCGGAGAGCCCATGGGTCAACCAATG 1177
Qy      1010 agtgcctcttaagaagaacatctttaagaacgagatataatgctgtaaatcac 1069
Db      1178 AATGGAAGAACGGGAGCGCAACTCTCAAGCCGATGTCTCTTACAAATCCAAATTC 1237
Qy      1070 aacattactacatggtatccaataatcgtcttaacagcagaatagagatacaagcta 1129
Db      1238 CGCATAGCTATATATTAACCTCTAATCAAAATCTGTCTGAAAAGCAAGTCAAAACCA 1297
Qy      1130 ctgttaaacatgcaaatggaagtggtgttaagtgaatgagctgcgtcttaattga 1189
Db      1298 CGGTAAAGCTAGCGAACTGGGCAATTCGCAACGGGATTTATGCAAACTATACGTA 1357
Qy      1190 aaacgagaaactgttatatactgtgactctggaagcttgccgagtaaaaggttaccctg 1249
Db      1358 AGACCGGGTCCGGGTGACATGTGTACATTCGAGAGAAATGTCAATCAACGACG 1417
Qy      1250 gaacacgcctatctagatttacaatgtccaataatctctcaagtaagggaacttg 1309
Db      1418 GTACCAATTTTGAACATACCTCCCTCAGCGCATTCGAAATTTGCTCCTCAAGAAATTTG 1477
Qy      1310 gaattcagttccaatcagcagtgatagtagtggacaacaatcgattatattgaaatg 1369
Db      1478 GGGTAAATTCGCCGCTCCTCAACAGTGTGCGCAATACGATATTTATGTATGTAGT 1537
Qy      1370 tgaattgaa 1380
Db      1538 TTAGTCTGCA 1548

RESULT 2
LOCUS   D86329 1530 bp DNA BCT 19-MAR-1998
DEFINITION Vibrio sp. gene for beta-1,4-mannanase, complete cds.
ACCESSION D86329.1 GI:2978250
VERSION beta-1,4-mannanase; manA.
KEYWORDS Vibrio sp. (strain:MA-138) DNA.
SOURCE Vibrio sp.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 1530)
AUTHORS Tamaru, Y.
TITLE Direct Submision
JOURNAL Submitted (01-JUL-1996) to the DDBJ/EMBL/Genbank databases. Yutaka
Tamaru, Mie University, Faculty of Bioresources, Utilization of
Marine Products, 1515 Kamihama, Tsu, Mie 514, Japan
(E-mail: s9tamay@maric.bio.mie-u.ac.jp, Tel: 0592-31-9561,
Fax: 0592-31-9557)
2 (sites)
REFERENCE Tamaru, Y., Arai, T., Morishita, T., Kimura, T., Sakka, K. and
Ohmura, K.
TITLE Cloning, DNA sequencing, and expression of the beta-1,4-mannanase
JOURNAL gene from a marine bacterium, Vibrio sp. strain MA-138
FEATURES
location/Qualifiers
1..1530

```

```

/organism="Vibrio sp."
/strain="MA-138"
/db_xref="taxon:678"
32..37
-35_signal
-10_signal
RBS
gene
CDS
136..141
149..1336
/gene="manA"
149..1336
/gene="manA"
/codon_start=1
/translation="beta-1,4-mannanase"
/protein_id="BA2518.1"
/db_xref="GI:2978251"
/translation="MKETAFISLLEIFVASCAGHAEFYVNGVLFEANGSAPRKING
HAHTWYTDKLSVALSGIAATGANTYAVVLSNGRWKNDVSDVTNINAKANLAI
LEVHDTGGESSASLSDSADYVTELKNEILGODEYVILNGNPFENNDAVAV
NDHVSAIOLRBSAGINHTIMVDAPNMGDMPLNAGFVNSDKLMTIFSVHME
VYSYNSVDYISSFTNNGLVLYIGFPAETHKGADYDESIMRSETLSIGYGSWS
GNDTTSDDIDIVNNMDNNSYSTMGNVLYIGONGIKSTLYLFTFTGNDCCDSCEY
PICSSAVDPDDGDGWMENNOSCIVODSSDITAPNGVPCYQSSSDPDGDGWMENNAS
MCC"

BASE COUNT 469 a 211 c 333 g 517 t
ORIGIN

Query Match 17.5%; Score 242; DB 2; Length 1530;
Best Local Similarity 57.2%; Pred. No. 4, 9e-48;
Matches 525; Conservative 0; Mismatches 375; Indels 18; Gaps 4;

Qy      3 tgcgaattccggaatttatgaagcgggtaccactctataagaatgcaatgaaaccatt 62
Db      202 TGCTCATGCGGGTATTATATATCAATGCTGTATATGAAGCGAATGTGTGCTT 261
Qy      63 tgaatgagaggattaaaccatggcagcatggtataagaacagcaactactgcaat 122
Db      262 TAAATTTAGGGGATTAATATATGACACTTGTGTATACGACAAAGTTAAGCTGCTT 321
Qy      123 tgaaggattgcaataacccggtgcttaatacgcgtcgatgtgttactatgaggggaca 162
Db      322 AAGTGTATTGCGGCCACAGACGACCAATAGTACGGGTGTACTAAGTAAAGTATCG 381
Qy      183 atggacaagaatgacatccatcagtaagaacctatctctttagcggaaagataatc 212
Db      382 GTGACAAAAAATGACCTGTGATGTGACTAATTAATTAATTAACAAACCAACAA 441
Qy      243 ttggtgtcgtctctgaagtatgatgactgacggtatg-----attccattg 293
Db      442 CTTTATAGCTAATTTAGAACTTATGACACTTACTGGGTATGGAGAAAGTATGCTGTC 501
Qy      294 ttgcctcaatcgtcgtctgtattattatgattgaatgagaagtgctttaaattgaaaga 353
Db      502 TAGTTTATGATTCGGCGGCTATTCAGTATGATGATGATGATGATGATGATGATGATG 561
Qy      354 agataccgtcatatataatcttgcgaatgaatgtgttg---ttcgtgggaaggagatgc 410
Db      562 AGATTATGTCATCATTAACCTAGCAATGAACTTTTGGTAATTAATTAATATCGGTAGC 621
Qy      411 ttggcgtcagcgggtataaacaagaacatcccgcgatctcgtaagcggcgcttaaccatc 470
Db      622 TTGGGTTAATGACCAATGTCTGCTATTCACGATTAAGAAGTGCAGATTCATTCACAC 681
Qy      471 ctgatgtgataatgctgctggtggtgggacaa---tttccaatcgaatcattatggtg 527
Db      682 GATTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 741
Qy      528 aagagaagtttttaagctgacccctcaagaataacaaatgcttgcattatgattatga 587
Db      742 TCAATTCGTTTAAATTCACCTCAATTAATGAATGATGATGATGATGATGATGATGATG 801
Qy      588 atatgcagtggtgaatgacgcgaatgtctgactaatattgaccaggttcttaacaga 647

```

Db	802	GGTTATATAGTAGTTACAANTC---AGTCATAGCATACATCTCATCTTTACAAACAATAG	858
Qy	648	ccctgcattatgcatcattggtgaatttggacacgcgtcatacaatggtgcgtgaagaac	707
Db	859	CTTAGGCTGGGTTATACGGGAGGTTCCTTCAACTATAAGGGGCCGAGTGGATGAAG	918
Qy	708	aacgatattgagctatttcgaaacaagaggttgggtgttggcggtgtcatggaag	767
Db	919	TTCAATTAATGAACGTTTCAGAAACGTTAAGTTAGTTATTCGCGTGGTCTCGTCAAG	978
Qy	768	gaacgccccagaatgggagatttagaccttcgaatgattggtcgtgaaataacctac	827
Db	979	AAATGATTAATACACATCCGATTTGGATTGTTGTTATTAATTTGGGACAATAATTCGATAG	1038
Qy	828	agcttgggggaatacaatgtagatggtcatatggtttaaagaagaactcgaattaa	887
Db	1039	CACTTGGGTAATGTTTATTAATTAATGTCAAATGATTAATTAATCCACATGACGTTAGC	1098
Qy	888	cacgctttttacaggtg	905
Db	1099	TACGGTGTTTACTGTGG	1116
RESULT 3			
AF163837..	LOCUS	AF163837	4567 bp DNA BCT 08-FEB-2000
DEFINITION	Caldibacillus cellulovorans multidomain beta-1,4-mannanase precursor (mana) gene, complete cds; and unknown genes.		
ACCESSION	AF163837		
VERSION	AF163837.1 GI:6651325		
KEYWORDS			
SOURCE	Caldibacillus cellulovorans.		
ORGANISM	Caldibacillus cellulovorans.		
	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Caldibacillus.		
REFERENCE	Sunna, A., Gibbs, M.D., Chin, C. W. J., Nelson, P. J. and Bergquist, P. L. A gene encoding a novel multidomain beta-1,4-mannanase from Caldibacillus cellulovorans and action of the recombinant enzyme on kraft pulp		
AUTHORS	Appl. Environ. Microbiol. 66 (2), 664-670 (2000)		
TITLE	2 (bases 1 to 4567)		
JOURNAL	Sunna, A., Gibbs, M.D. and Bergquist, P. L.		
MEDLINE	Submitted (29-JUN-1999) Biological Sciences, Macquarie University, North Ryde, New South Wales 2109, Australia		
REFERENCE	location/Qualifiers		
FEATURES	1..4567		
source	/organism="Caldibacillus cellulovorans"		
	/db_xref="taxon:74586"		
	<1..513		
CDS	/note="orf1"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AA022273.1"		
	/db_xref="GI:6651326"		
	/translation="TPTPTPSAAPPTPSAGSLVYQYRAADTNAGNQLAPHERI		
	VNGTSVPSLELTIWYVDGKPEVFCMDVAQVCSNVRGSEFVLSTGRGADY		
	IEITFSAGSLAAGASSGDIQVIRINNDWTINNEANDYSIDPTKSFADMNRYTLR		
	NGDLINQVEP"		
	1..60		
misc_feature	/note="Region: proline-threonine linker"		
	/evidence=not_experimental		
	61..510		
misc_feature	/note="Region: cellulose-binding domain type IIIB"		
	/evidence=not_experimental		
	620..3412		
gene	/gene="mana"		
	620..718		
sig_peptide	/gene="mana"		
	/evidence=not_experimental		

[illegible]

OY	128	ggattgcaaatcccggtctctaataagcgcggaatgtgttacttgatgaggacaatga	187		
Db	2130	GCATTGTTCTCGGGCCGCCAACGCGGTGGATGTGTGTAGCAACGGTTCGGGTGA	2189		
OY	188	caaaagaatgacatccatacagtaagaaaccttatctcttaacggyaagataa---tcatt	244		
Db	2190	GGAAGATTCCGGCAGCGAAGTAGGAGCATATTATTCTCAGCGCGGAGACGTCGGTTATTC	2249		
OY	245	tgtgtcgtgtctctttaaagtcatgatgtcaccgttgtatgatcattgc-----tt	295		
Db	2250	GGGCGGTGTTCTTAATAATTCCAGCACAGACCAGCGGTACGGGGAMACGCGGGCGGTGT	2309		
OY	296	cgcctaacctgcgtgttgatatgttgatgaaatgaagatgcctttaattgaagaagag	355		
Db	2310	CGATACGACGCGCGGTCAATTATTGGATTGAACCTAGAAGACTGCTGGCGGGTCAAGACA	2369		
OY	356	atacgcctcataattaataattgcgaatgaatggttttggttcgtggyaagaagagatgctgg	415		
Db	2370	ATTTCGTATGTCATTAATTACGGCATATGAGCCGTACGGGAAACAATTATTATCAAGACTGGG	2429		
OY	416	ctgaegygatataaacaagaatcccgcgagttgcgtaaogcgggtctaaccattctga	475		
Db	2430	TGACGACACCGCGGAACCGGTGCAGGGCGCTCCGMAACGCGGGATCAACAATACGATCA	2489		
OY	476	tgtgtatgctgcgaggggtggggac---aatcccaaatcgatltatgtatgtgaagag	532		
Db	2490	TGTGGAGCGTCCGAACCTGGGGTCAAGACTGCTCTTACGATGCGGGGCHAAACGCGCGA	2549		
OY	533	aagttttaatgctgtaccctccaacgaataatacatgttttcsgatcatalgtatgatatg	592		
Db	2550	CGATTTTCAACGCGGAGCCCGAGAGGAACCTGCTTTTTCATTCACATGTAAGGG---G	2606		
OY	593	caggtgtgtaatgcatcgcaagttcgtaactabatatgtgaccgagttccttaacaagacctg	652		
Db	2607	TGTACGACACGCGCGGGAAGTGCAAAGTTAATTAGAGCTGTTCTTAAACCCCGGGGTTCG	2666		
OY	653	catgtcattgtgtgtaatttgagcacgcgtcatatacaaatgtgtgacgtcgataagaacaag	712		
Db	2667	CGCTGTGATCGBGGAAATTGGTCTAATGACACGACGCGGGATCCGAACGAACGCGCA	2726		
OY	713	ttatgtagctatcttgaaacaaaagagagttggtggtgttgctgctgcatgpyaaagaagacg	772		
Db	2727	TGCTACATTATTCGAAGCAATACAACTATGGCTTGTGGCGTGTGTCGTGAGACGGTAACG	2786		
OY	773	gcccggaatggagagtcatttagacccttcgaaatgatgtggcgtgysaataaaccttaacgctt	832		
Db	2787	GAGCGCGGTGGAAATTGTGGACATGTGACACAACTTCATGTCGAACACCCGACGCGGT	2846		
OY	833	ggggaaa 838 Db 2847 GGGGGA 2852			
RESULT 4 CDCMANABD					
LOCUS	CDCMANABD	2155 bp	DNA		
DEFINITION	C. saccharolyticum beta-D-mannanase (mana) gene, complete cds.				
ACCESSION	M36063 M37147				
VERSION	M36063.1 GI:144292				
KEYWORDS	1,4-beta-D-mannan mannanohydrolase; beta-D-mannanase.				
SOURCE	C.saccharolyticum DNA, clone lambda-nzp2.				
ORGANISM	Bacterium Firmicutes; Bacillus/Clostridium group; Bacteriota; Firmicutes; Bacillus/Clostridium group; Thermanaerobacter group; Caldicellulosiruptor.				
REFERENCE	1 (bases 1 to 2155)				
AUTHORS	Luehl,E., Jasmel,N.B., Grayling,R.A., Love,D.R. and Bergquist,P.T.				
TITLE	Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium Caldocellulum saccharolyticum				
JOURNAL MEDLINE	Appl. Environ. Microbiol. 57, 694-700 (1991)				
COMMENT	Draft entry and computer-readable sequence kindly submitted by				

E. lueethi, 03-JUL-1990.
Author address: E.lueeth
University of Auckland
Department of Cellular and Molecular Biology
Private Bag
Auckland
NEW ZEALAND
email: SYSTEM@CEL.VAX.AUKUNI.AC.NZ.

FEATURES

source Location/Ovaliflers
.2155
/organism="Caldicellulosiruptor saccharolyticus"
/db_xref="taxon:44001"
<1..685
/codon_start=2
/transl_table=11
/label=ORF1
/protein_id="AAU72860.1"
/db_xref="GI:144293"
/translation="GSMTWFGFOAWSMQRYAEYYVYTGDKDAGALLKRWISWTIKSYWG
LNSDGTFAIPSTLDMSGPPPTWNNTYGNNRNLHAKYVDYTGDLTASLANALIYSN
GTKRVGYFEDEANKLNALKLELDRMKLTRDGKSAPKRADYKRFEEQEVIIPAQMIA
KMPGVDIKSGVFIDIRSKYKODPDPMFKLEAAKYSCQVEEFRHYRMACODIATIVNA
TYELIFGNQ"
TVELFNQ
818..830
/note="mana ribosome binding site"
841..963
/product="beta-D-mannanase"
841..1881
/EC_number="3.2.1.78"
/codon_start=1
/transl_table=11
/product="beta-D-mannanase"
/protein_id="AAA72861.1"
/db_xref="GI:144294"
/translation="MRLTKRKIRKWLVLCTVPELIITFIANTYLPRKYGAATSNDG
LVRIEDTLIGTNHGWCRDLDTLAAGRSMGMNVPLVLSGRWRTKIASEVAN
TISRSILGFRAILLIEVDHTTGEGEDAASSLAQAVFYWEIKSLVGNDEFVING
NEPGLNNNYNMWDNTAIKALIDADEFKITIMVDAENWCODMSNTNRDNQSIEMAD
PLRLVFSIHMYGVYNTPASKEVEEKSFVDKGJLVIGEFGHOTDDDPDEAIRVA
KOKRIGLFSSMCNSSYVGVIDMVNMMDPNPTPGOMYKTNAIGTSIPTPSVT
PTPPRRGHSHRO"

mat_peptide 964..1878

BASE COUNT 721 a 334 c 590 g 510 t

ORIGIN

Query Match 9.6%; Score 131.8; DB 2; Length 2155;
Best Local Similarity 51.7%; Pred. No. 1.8e-21;

Matches 407; Conservative 0; Mismatches 362; Indels 18; Gaps 4;

Dy 68 tgagagggaataaacacggygcagcagcatgtttaagaaccagcgcaactactctaatgaa 127
Db 998 TAAATGGACAACAATCACCCACCATTGCCTGGTACAGAGAATGACTTGATCGGCATTGCCGTG 1057

Oy 128 ggattgcaaataccggtytctaataccgtycccgatgttgttaacctgatggyggacaatgaa 187
Db 1058 GAATTAGGTCAATGGGGTATGGAACCTCTGGAGGGTAGTCCTGAGTAATGGTTAACCGATGGA 1117

Oy 188 caaaaagtacatcacatcatacgaataagaasacttatcccttgtagcggaagataatcattgg 247
Db 1118 CGAAAGATACCAACAAGTCAAGTGGCCAATAATTATATCATCTTCAAAGAAGCTTGGTTTCA 1177

Oy 248 ttgttgtt-----ctgaagttcatgattgttacccgittgatatgattcattg-----ctt 295
Db 1178 AAGCTATTATATTGAAAGTGCACACACAMACAGATATAGGAAGAAGTGGCCACCATGTT 1237

Oy 296 cgctcaatcsgycgcttgatttcattgattgaaatgaagaagtccttaatttgaaagaag 355
Db 1238 CATTTGGCACAGCAAGTGGAAATATGGAAGGAGATPAAGAGCGCTATTAGACGTAACGAAG 1297

Oy 356 ataccgctcatattaataattgcgaatgaaigtgttgcttgagggaagggaatgcctggg 415

Oy	773	gccccgaatggagcatttagaccttcgcaatgatgttggcctgaaataactacagctt	832
Dd	1715	CGAGCTATGTTGGTATTGTGCACATGTAAMACACTGGGACCACAATATCATTCAT	1774
Oy	833	ggggaaa	839
Dd	1775	GGGGACA	1781
RESULT	6		
CSU16308			
LOCUS			
DEFINITION	CSU16308	5284 bp	DNA
ACCESSION	Caldocellum saccharolyticum endoglucanase/mannanase (celC)		BCT 10-FEB-1996
VERSION	pseudogene, complete cds.		
KEYWORDS	U16308.1	GI:577827	
SOURCE			
ORGANISM	Caldocellum saccharolyticum. Caldicellulosiruptor saccharolyticus Bacteria; Firmicutes; Bacillus/Clostridium group; Thermoanaerobacter group; Caldicellulosiruptor. 1 (bases 279 to 4374) Morris,D.D., Reeves,R.A., Gibbs,M.D., Saul,D.J. and Bergquist,P.L. Correction of the beta-mannanase domain of the celC pseudogene from Caldicellulosiruptor saccharolyticus and activity of the gene product on Kraft pulp Appl. Environ. Microbiol. 61 (6), 2262-2269 (1995) 95314280 2 (bases 1 to 5284) Morris,D.D. Direct Submission Submitted (25-OCT-1994) Daniel D. Morris, School of Biological Sciences, University of Auckland, Private Bag, Auckland 92019, New Zealand		
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
FEATURES			
source	location/qualifiers		
	1..5284		
	/organism="Caldicellulosiruptor saccharolyticus"		
	/db_xref="taxon:44001"		
	279..341		
misc_feature	/gene="celC"		
	/note="encodes leader sequence"		
gene	279..4374		
	/gene="celC"		
CDS	279..4374		
	/gene="celC"		
	/note="The celC gene potentially encodes a multidomain enzyme comprised of an N-terminal endoglucanase domain, two central cellulose-binding domains, and a C-terminal mannanase domain. However, due to the presence of two open-reading frame shifts in the mannanase-encoding region of celC, the actual enzyme is truncated and comprises only the endoglucanase and cellulose-binding domains"		
	/pseudo		
	/product="endoglucanase"		
	279..3572		
	/gene="celC"		
	/note="ORF #1 from celC start codon to frameshift #1"		
	342..1715		
misc_feature	/gene="celC"		
	/note="encodes endoglucanase domain"		
	1722..2204		
misc_feature	/gene="celC"		
	/note="encodes cellulose-binding domain (CBD) number 3"		
	2205..2342		
misc_feature	/gene="celC"		
	/note="encodes proline-threonine linker number 1"		
	2343..2807		
misc_feature	/gene="celC"		
	/note="encodes cellulose-binding domain (CBD) number 1"		
	2808..2957		
misc_feature	/gene="celC"		

	misc_feature	/note="encodes proline-threonine linker number 2"
	2958..3419	/gene:"celc"
	/note="encodes cellulose-binding domain (CBD) number 2"	
	3420..>3524	/gene:"celc"
	/note="encodes partial mannanase domain"	
	3420..3524	/gene:"celc"
	/note="encodes proline-threonine linker number 3"	
	3574	/gene:"celc"
	/note="location of the first -1 frameshift"	
	3575..3625	/gene:"celc"
	/note="ORF #2 from frameshift #1 to frameshift #2"	
	3627	/gene:"celc"
	/note="location of the second -1 frameshift"	
	3628..4374	/gene:"celc"
	/note="ORF #3 from frameshift #2 to celc stop codon"	
	BASE COUNT	1829 a 881 c 1238 g 1336 t
	ORIGIN	
Query Match	9.1%; Score 125; DB 2; Length 5284;	
Best Local Similarity	51.1%; Pred. No. 7.6e-20;	
Matches	405; Conservative	0; Mismatches 370; Indels 18; Gaps 4;
QY	121 attgaaggattcgcaaataccggtgcctaatacgcgcggatgttctctcyatggsgaa	180
Db	3610 ATTGGTGCAGATTCAGCGCATGGGGGTATGAACTCTTGAGAGGGTAGTCAAGTAATGGTTAC	3669
QY	181 caatggaaaagaagatgatcatcatcaacaaagaaacctatcccttttagcggaagatat	240
Db	3670 CGATGGACGAAGAATGCCAGCAGATGAAGTGGCAATTATTAATTCATTTGCAAGAGCTTT	3729
QY	241 catlgtgttcgttcttctgaagtcatatgcttacgcgttatgatccatg-----	292
Db	3730 GGTTCAGAACGATTATATTAGAGTACACGACAACAGGATATGGAAGATGGGCA	3789
QY	293 ----cttcgcctaactcgtctgttgtatatttgattggaatgagaagtgctttaattgaa	348
Db	3790 GCGTGTTCATTTGTCACAGCAGTGGAAATATGGAAGAGATAAAGAGGTATTAGACGT	3849
QY	349 aaagaagataccgcattatataatattbgcgatgatattgttgcgttggaaagggat	406
Db	3850 AACGAAAGTTTTGTATTTATTAACATTGGCAATGCCGTATGGCAACATAACTATCA	3909
QY	409 gcttggagctgcgggtataacaagaaccaatcccgcgattcgtaacgcgcgtctaaccat	468
Db	3910 AACTGGCTTATGACACGCAAAACGCTTTAAAAGCACCTTAAGATGCGAGATTCAAGCAC	3969
QY	469 accttgaatgtagatgctgcggggtyggagac--aatttcacaactgattcatat	525
Db	3970 ACGATTAATGGTGGATGCGCCAGACGGGGGTACAGATTGGTCTAATATCATGAGATTAAT	4025
QY	526 ggaagagaagattttaatgctgcgccctcaagaagaataaattgcttgcattcatatgcat	565
Db	4030 GCCCAGACCAATAAGGAAGCAGATCCCCTCGGCATTTGGTATTTTCATTCAATATATAT	4089
QY	586 gaatatgagtggttaatgacatcgcaagtgctactaatattgaccgcggtcttaataa	645
Db	4090 G---GTGATATCAATATACGCAAGCATGTCAGAGATCATCAATCATTTTGTTATAG	4146
QY	646 gacctcgcatatgcatatgctgtaatttggacaacgcttacaanaatlygtaagctgcatgaa	705
Db	4147 GGGTATACCATTTGGTTATTTGGGGAATTTGGACATCAGCACACAGATGGTGAACCCGATGAA	4206
QY	706 gcaacgattatgagctattctgaaacaagaaggagtggtggtggtggtggtgcatgtaa	765
Db	4207 GAACCTATTGTCAGATTTGCAAAACAGTACAGATAGCATTAATTTAGTTGGTCGGTGT	4266

QY 766 gggaaacggcccaagatgagatatttagaccttcgaatgattggctggaataaccc 825
 DB 4267 GGAATTCGAGCTATGTTGGGTAATTGACATGTAACAACATCGACCCCAATATCA 4326
 QY 826 acagcttgggggaatacaataagatgagtcacataatggtttaagaagaacttcgaat 885
 DB 4327 ACTCCATGGGGGCAATGTAATAAATGCAATGTAACAAAGATGAATAAATTA 4386
 QY 886 agacacgtttta 898
 DB 4387 TTCAAAGTCTTA 4399
 RESULT 7
 LOCUS E02075 1461 bp DNA PAT 29-SEP-1997
 DEFINITION Genomic DNA encoding beta-mannanase of Alkalophilic Bacillus
 sp. AM-001.
 ACCESSION E02075
 VERSION E02075.1 GI:2170317
 KEYWORDS JP 1989228477-A/1.
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 1461)
 AUTHORS Akino, T. and Horikoshi, K.
 TITLE BETA-MANNANASE GENE DNA, RECOMBINANT PLASMID CONTAINING SAID DNA AND
 TRANSFORMANT
 JOURNAL Patent: JP 1989228477-A 1 12-SEP-1989;
 RES DEV CORP OF JAPAN, AKINO TOSHIRO, HORIKOSHI KOKI
 COMMENT OS Alkalophilic Bacillus sp.
 PN JP 1989228477-A/1
 PD 12-SEP-1989
 PF 09-MAR-1988 JP 1988053774
 PI AKINO TOSHIRO, HORIKOSHI KOKI
 PC C12N15/00, C12N1/20//C12N9/24, (C12N15/00, C12N1/07), (C12N1/20,
 PC C12N1/19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC key *source: strain-AM-001;
 FH Location/Qualifiers
 FT CDS 1..1461
 /product="beta-mannanase of Alkalophilic Bacillus sp."
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 475 a 257 c 340 g 389 t
 ORIGIN
 Query Match 7.9%; Score 109.4; DB 5; Length 1461;
 Best Local Similarity 55.0%; Pred. No. 4,4e-16;
 Matches 259; Conservative 0; Mismatches 26; Indels 6; Gaps 2;

DB 1168 ACAGTAATCGAATATGTTGATTCATTAACAAACCTGAAATTAATGACCCATTTCG 1227
 QY 1147 tggggaagtcgttgtaataagtaactgcgccttattatggaacagagactggtat 1206
 DB 1228 TGGGGAATTCGAGAGTGGCATGACAGCAAGTTTGTGCAAAACAGGAACTGGTGG 1287
 QY 1207 acatgtaactctggaagcttgcgcgattacacggttcattcgtgaacacgctatctta 1266
 DB 1288 AGATGGAATCGAGTGAATTTGTGAGTTGCGAGCAACGAACCAACCGCATATCTATT 1347
 QY 1267 gattatcaatgtccaaactcttcacagtaagggaattgagttcagttccaatca 1326
 DB 1348 GATTGACGAAGTAAGTAATTCGATGATGTCGAGACATAGCTAGCTTAAGCA 1407
 QY 1327 gcgagtatagtatgacacaacatcattatattatgataatgattgta 1377
 DB 1408 CCAGCAAAATGCAACGCGAAGACGGGATTTACTTATGATCATGTGACGGTA 1458
 RESULT 8
 LOCUS BACMANN 1939 bp DNA BCT 26-APR-1993
 DEFINITION Bacillus sp. beta-mannanase gene, complete cds.
 ACCESSION M31797
 VERSION M31797.1 GI:143166
 KEYWORDS 1,4-beta-D-mannan mannanohydrolase; mannan
 endo-1,4-beta-mannosidase; mannanase.
 SOURCE Bacillus sp. (strain AM-001) DNA, clone pMAN5.
 ORGANISM Bacillus sp.
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus.
 REFERENCE 1 (bases 1 to 1939)
 AUTHORS Akino, T., Kato, C. and Horikoshi, K.
 TITLE Two bacillus beta-mannanases having different COOH terminal are
 produced in Escherichia coli carrying pMAN5
 JOURNAL Appl. Environ. Microbiol. 55, 3176-3183 (1989)
 MEDLINE 90146329
 FEATURES
 source Location/Qualifiers
 sig_peptide 1..1939
 /organism="Bacillus sp."
 /db_xref="taxon:1409"
 220..297
 /note="beta-mannanase signal peptide"
 CDS 220..1761
 /note="beta-mannanase precursor (EC 3.2.1.78)"
 /codon_start=1
 /transl_table=11
 /protein_id="AA02586.1"
 /db_xref="GI:143167"
 /translation="MKYKKVAFVAFIMFVSVPITISMSSEANGAALSNPNANOTTK
 NYSWLANLPNKSRRVSGHFGSDSTLMTKOCARELTGMPDILSCDYKNQTP
 LVVADDISGNOELINFNQGLVTFIRHPNPGHSENTKTLTISQFQNLTPR
 TTEGRKMDLMDKMDGLDELQNGTVLFRPLHENGEMFWMGAGYNOFQDRWA
 YISAWMDQYFTHERKLNLLWVYSPDYRDVTSYYPGAVDIVALDSYHPDHS
 LIDQVYRMALDKPFAFATIGPESWAGSEDSNYTOAIKOKYPRVYVLANDKSP
 HNRKANDLFDNSWYVNRGEIDYGSNRPATVLYDPENNLTLSGCCFTGCGPESHA
 SANGTSLRADVYLGNNSTYHLQTVNRNLSSENRKIKSHSNGVGSGLKRVYLA
 TSAGAWNRNGEFCQFAGKRTALSIDLTVYSNLDHREIGVEYKAPANSNGKALITL
 HLYVR"
 mat_peptide 298..1758
 /note="beta-mannanase"
 misc_signal 1826..1840
 /note="pot. transcription termination signal; putative"
 misc_signal 1861..1889
 /note="pot. transcription termination signal; putative"
 misc_signal 1905..1916
 /note="pot. transcription termination signal; putative"
 BASE COUNT 637 a 343 c 420 g 539 t
 ORIGIN 1 bp upstream of XbaI site.
 Query Match 7.9%; Score 109.4; DB 2; Length 1939;


```
Db 25173 ACGGCCGACCTCCGCGGCGGCTACCTCGCTGCTGCTGAGAGGCGGCGGCGGC 25232
QY 781 tggagatttaagacccttcgaatgatggcctggaataaacttaacgttgggaat 840
Db 25233 GTCGAGTACCTCGACAGTGTGACCGGCTTCCAGCCGAACTCGACGAGGTGGGGCAAC 25292
QY 841 acaatagtaagtgcataatgatttaagagaacttcgaatgaatgaacgttttaca 900
Db 25293 CGCATCTTCTACGCGACGACGATCGCCGCGACGCTCCAGAGCGCCACCGTGTACGCG 25352
QY 901 ggtggaag 908
Db 25353 GGTGTGTG 25360

RESULT 10
STMBMANASE 1881 bp DNA BCT 16-JUN-1999
LOCUS STMBMANASE Streptomyces lividans mananase (mana) gene, complete cds.
DEFINITION M82297
ACCESSION M82297.2 GI:4579683
VERSION M82297.2
KEYWORDS Streptomyces lividans
SOURCE Actinobacteria; Actinobacteriales; Streptomyces
ORGANISM Bacteria; Firmicutes; Actinomycetales; Streptomyces.
REFERENCE 1 (bases 1 to 1881)
AUTHORS Arcand,N., Kluepfel,D., Paradis,F.W., Morosoli,R. and Shareck,F.
TITLE Beta-mannanase of Streptomyces lividans 66: cloning and DNA
JOURNAL sequence of the mana gene and characterization of the enzyme
MEDLINE Blochem. J. 290, 857-863 (1993)
93207541
REFERENCE 2 (bases 1 to 1881)
AUTHORS Shareck,F.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1992) Centre de recherche en microbiologie
applied, Institut Armand-Frappier, 531 boul. des Prairies, Laval,
Quebec H7N 4Z3, Canada
3 (bases 1 to 1881)
REFERENCE 3 (bases 1 to 1881)
AUTHORS Shareck,F.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1999) Centre de recherche en microbiologie
applied, Institut Armand-Frappier, 531 boul. des Prairies, Laval,
Quebec H7N 4Z3, Canada
REMARK Sequence update by submitter
COMMENT On Apr 8, 1999 this sequence version replaced gi:153193.
FEATURES
source
1..1881
/location/Qualifiers
1..1881
/organism="Streptomyces lividans"
/strain="1326"
/db_xref="taxon:1916"
217..1316
/gene="mana"
217..221
/gene="mana"
226..1377
/feature="putative"
226..1377
/gene="mana"
/EC_number="3.2.1.78"
/codon_start=1
/translation="11"
/protein_id="AAA26710.2"
/db_xref="GI:4579684"
/translation="MNRARSTLTITAGMAFVILGLFALAPSGRAEAAGIHYNS
GRAVEGSAFVKGVNHATYTPDRTGSIAIAKAGANTVRYVLSGGHWTTSASE
VSAIIGCKKANKYICVLEVDHTGYGKGATSIDAGADYVYKSAAMRQEDYVYV
IGNEPENTNYAATDATKSAIGLRAGSGHALMDVAPMGQDSSTMAASVTA
SDPDRNTVESIHMTGVTDALEVDYINAVVNGCLPVEFEQHSRDNDEDAIVA
TAOSLAVGLMGMSGNGGVEYIDMNVNGDPNLSISWGRNIIYSGNSGIASTRTAV
YGGGGSGTGTABNGPYCYVNGASDEPDGDGMEWENRSRCSVVRGSAADH"
226..333
sig_peptide
```

```
mat_peptide /gene="mana"
334..1374
/EC_number="3.2.1.78"
/function="hydrolase"
/product="mannanase"
BASE COUNT 309 a 648 c 647 g 277 t
ORIGIN
Query Match 6.5%; Score 89.2; DB 74; Length 1881;
Best Local Similarity 47.6%; Pred. No. 3.2e-11;
Matches 432; Conservative 0; Mismatches 458; Indels 18; Gaps 5;
QY 13 ggaatttatgaagcgggtaccacactctatagatgcaatggaacccattgtaatga 72
Db 340 GGCATCCAGCTCGACACGACGAGTGTCTCGAGGCAACGCGACGCGTGTCTATGCGC 399
QY 73 ggaattacatggaagcgcgcgtatataaagacgaagcaactctgaattgaaggaatt 132
Db 400 GGGGCAACACGCGCTACACTGTGATCCGACG---GCACGGGCTCATGCGGACATC 456
QY 133 gcaaataccggtgctaatacagtcgcggaattgtatctatgattggyggaacaa 192
Db 457 GCGGCCAAGGAGGCAACACCGCTCGCTCGCTCAACAGCGCGCGCTGACAGAG 516
QY 193 gatgaatccctaaagtaagaacactatctcttaagsgaagataatcattggtgt 252
Db 517 ACGAGCGGCTCGAGAGTCTCGGCTCTATCGGCAATGCAAGGCCAACAGTATCTGT 576
QY 253 gtctctgaagtcataatgatacgcggtatattccattgcttcgca-----atcgt 306
Db 577 GTCCTTAGAGTGCACAGACACACCGGCTACGCGCAAGAGCGGCGGACCGCTGAGCAG 636
QY 307 gctgttgattatgattgaatgaagatgct---ttaatggaaggaagatacgcgtc 363
Db 637 GCGGCGGACACTGAGTGGTGTGTGAAGAGCGCCGCTGAGAGGCGGAGGACTACGTC 696
QY 364 attattaatatgtaagatgattggtgttcgtyggaagsgatgctgtgcyggaaggg 423
Db 697 GTCGTCAACATCGGACAGAACCCCTTCGGCAACCAACCACTACGCGCGTGAACCGG 756
QY 424 tataacaagaatccgcgcgtatgctgaacgcggtctaaacatacattgattgatt 483
Db 757 ACGAAGTGGCGCATCGGACGACACTCGCGGCGCGGTCTCGGCGACGCGCTATGTCGAC 816
QY 484 gctcgcggggtgggac---aattccaacatcgattatgattggaagaaagatttt 540
Db 817 GCGGCCAAGTGGGCGCAGAGACTGTCGCGGACGATGCGGTCAACGCGGCGCTCGTTC 876
QY 541 aatgtgacctcaagaataataaatgttctgattatataatgattatgacggtgt 600
Db 877 GCGTCCGACCGCGACCGACACCGCTCTTCATCCACACATGACG---GGGTACGAC 933
QY 601 aatgcatcgcaagtcgtactaatattgacccgaggtctttaaagaacccctgcatatgc 560
Db 934 ACCCGCGCGAGTGCAGGACGATCACTTAACCGCTTCGTCGCGCAACGACATCCATGTC 993
QY 661 attgtaatttgcacacgcgtcatacaatattgtaacgttgcgtgaagaacgattatgac 720
Db 994 GTCGGCGAGTTCGGCGACACACAGTACGCGCAACCCCGGAGGAGCGCATCATGTCG 1053
QY 721 tatctgaacaagaagagattggtgtgttcgtygctcatgaaaggaagggcccgaa 780
Db 1054 ACGGCCGACGCTCCGCGCTGCTGCTACCTCGCTGCTGCTGAGCGGCAAGGGGCGGC 1113
QY 781 tggagatttaagacccttcgaatgatggcctggaataaacttaacgttgggaat 840
Db 1114 GTCGAGTACCTCGACAGTGTGACCGGCTTCCAGCCGAACTCGACGAGGTGGGGCAAC 1173
QY 841 acaatagtaagtgcataatgatttaagagaacttcgaatgaatgaacccgttttaca 900
Db 1174 CGCATCTTCTACGCGACGACGATCGCGCGACGCTCCAGAGCGCGCACCGTGTACGCG 1233
```


OY 901 ggtgagag 908
 |||||
 Db 1234 GGTGTGG 1241

RESULT 11
 LOCUS TFU6227 837 bp DNA
 DEFINITION Thermomonospora fusca man gene, partial. BCT 03-DEC-1998
 ACCESSION AU006227.1
 VERSION AU006227.1 GI:3970819
 KEYWORDS beta-mannanase; man gene.
 SOURCE Thermobifida fusca.
 ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptosporangiales; Nocardiopsaceae; Thermobifida.

REFERENCE 1 (bases 1 to 837)
 AUTHORS Hilge, M.
 TITLE Direct Submission
 JOURNAL Submitted 13-MAY-1998 Hilge M., Biochemistry, Federal Institute of Technology Zurich, Universitaetstr. 16, CH-8092 Zurich, SWITZERLAND

REFERENCE 2 (bases 1 to 837)
 AUTHORS Hilge, M., Gloor, S., Winterhalter, K. and Piontek, K.
 TITLE Crystallization and preliminary crystallographic analysis of two beta-mannanase isoforms from Thermomonospora fusca KM3
 JOURNAL Acta Crystallogr. 52, 1224-1225 (1996)
 REFERENCE 3 (bases 1 to 837)
 AUTHORS Hilge, M., Gloor, S.M., Rypniewski, W., Sauer, O., Heighman, T.D., Zimmermann, W., Winterhalter, K. and Piontek, K.
 TITLE High-resolution native and complex structures of thermostable beta-mannanase from Thermomonospora fusca - substrate specificity in glycosyl hydrolase family 5
 JOURNAL Structure 6 (11), 1433-1444 (1998)

JOURNAL MEDLINE
 FEATURES
 SOURCE

1. 837 Location/Qualifiers
 /organism="Thermobifida fusca"
 /strain="KM3"
 /db_xref="taxon:2021"
 1. 837
 /gene="man"
 <1..>837
 /gene="man"
 /ec_number="3.2.1.78"
 /function="hydrolysis of mannan"
 /translation="GLHVKNGRLYEANGQEFIIINGVSHPNMYEQHTQAPADIKSHGA
 NTVRVVLSNGVRSKNGSPDVANVLSICQKRLICMEVHTTGYEGSGASTDQAV
 DYVIELKSVLGEDDYVILNIGNEPYGDSITVAGAMDTSALQIORLAAGFEHTLV
 DAPWGDWNTMKNADQVYASDPGTGVSISIMGVYSGASITSTYIEFVAGAP
 LIIGERHDSHSDGPDDEDTIAEAEKRLKLTIGMSNGSGGVYLDVNTNFDGDNLS
 PWG"
 /protein_id="CA06924.1"
 /db_xref="GI:3970820"
 /translation="GLHVKNGRLYEANGQEFIIINGVSHPNMYEQHTQAPADIKSHGA
 NTVRVVLSNGVRSKNGSPDVANVLSICQKRLICMEVHTTGYEGSGASTDQAV
 DYVIELKSVLGEDDYVILNIGNEPYGDSITVAGAMDTSALQIORLAAGFEHTLV
 DAPWGDWNTMKNADQVYASDPGTGVSISIMGVYSGASITSTYIEFVAGAP
 LIIGERHDSHSDGPDDEDTIAEAEKRLKLTIGMSNGSGGVYLDVNTNFDGDNLS
 PWG"

BASE COUNT 169 a 292 c 255 g 121 t

ORIGIN

Query Match 5.9%; Score 81.6; DB 74; Length 837;
 Best Local Similarity 48.0%; Pred. No. 2.2e-09;
 Matches 404; Conservative 0; Mismatches 414; Indels 24; Gaps 5;

OY 13 ggaattatgaagcggtaccactatcagatgcaatggaaccattgtaagaga 72
 |||||
 Db 1 GGGCGCATGTAAGAGAGCGCGCTGTAAGGCCAACGGCAGAGTTCAATCCGT 60
 |||||

OY 73 ggaattaacatggcagcagatgataaagccggaactactgcaattgaaggatt 132
 |||||

Db 61 GCGGTACGACACCCCAACTGCT---ACCCGAGACACACCGAGCTTCCGACATC 117
 OY 133 gcaataaccggtgtctaatacaggtccgattgtgtatctatctgaggaatgacaana 192
 |||||
 Db 118 AAGTCGACGGCGCCAACACCGTCCGGGTGTGTGACCAAGGTGTCGGTGAAGCAAG 177
 |||||

OY 193 gatgacatccatacagaagaacctatctcttaacggaataatcatattgttgt 252
 |||||
 Db 178 AACGCTCTTGTGACGTGCCAACGTATCTCCCTGTGCAAGCAACCGCTTATCTGCC 237
 |||||

OY 253 gttctgaagttcatatgctaccggtatgattccatt-----gttcgtccaat 303
 |||||
 Db 228 ATGCTGAGGTGACAGACACACCGGCTACGATGAGAGACAGCGGGCTCCACCTTGAC 297
 |||||

OY 304 cgtctgtgtgttttttgattgaaatgagaatgctcttaattgaaaggaataacgttc 363
 |||||
 Db 298 CAGCGCGTGCATCTGATGATGACTGAAGACGTGCTCCAGGGCGAGAGACTATGTC 357
 |||||

OY 364 attaatatctggaatgaaattgttgt-----tcgtggaagggatgcttggt 417
 |||||
 Db 358 CTCATCAACATCGGCAACGACCTTACGCAACATCCGACCTGCGCTGGGGCG 417
 |||||

OY 418 gacggtataacaagaacatcccgatgtgtaacgcccgttctaaacataacctgtatg 477
 |||||
 Db 418 TGGGACACCTCCGCGCATCCAGCGCTGCGCGCGGATTCGACACACCTCGTG 477
 |||||

OY 478 gtaatgtctggggtgtggac---aattccaaatcatgattatgaaagaga 534
 |||||
 Db 478 GTGAGCGCCCACTGGGGGCGAGACTGAGCAACCAATGCGGAACACCGGACAG 537
 |||||

OY 535 gttttaatgctgcccccaagaatacaatgatttctgattcatatgaaatgca 594
 |||||
 Db 538 GTGACGCCACGACGACCCACCGGCAACCTCTTCTGATCCACATGTAG---GGCTC 594
 |||||

OY 595 ggtgtaatgcatgcgaagtctgactaataattgacgagttcttaatacaagacctgca 654
 |||||
 Db 595 TACTGCCAGGCGTCCAGATCACAGATCCTGAGGACACTTCTGCAACGCGGCGCTGCG 654
 |||||

OY 655 ttatcatgtgtgattgtgacacgcgtcatcaaatgattgtgcgctgaaagcaagatt 714
 |||||
 Db 655 CTATCATCGGCGAGTTCGCGCACGACACTCCAGCGCAACCCCGAGAGACACGATC 714
 |||||

OY 715 atgaactatctgacaagaagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 774
 |||||
 Db 715 ATGCGCGAGCGCGAGCGGCTCAAGCTGAGCTACATCGGCTGTGTGTGTGTGTGTGTGT 774
 |||||

OY 775 ccagaatggagattttagaccttgcgaattgattggtgctggaataaccttaacgtttg 834
 |||||
 Db 775 GCGGGGCTGAGTACCTGACATGATGTGTAACTTCGACGCGCAACCTGAGCCCGTGG 834
 |||||

OY 835 gg 836
 |||||
 Db 835 GG 836

RESULT 12
 LOCUS SCF73 14922 bp DNA
 DEFINITION Streptomyces coelicolor cosmid F73. BCT 04-OCT-1999
 ACCESSION AL121746.1
 VERSION AL121746.1 GI:6013076
 KEYWORDS catalase/peroxidase; cpeb; Fe regulatory protein; furs; manA; membrane-bound oxidoreductase; Merr-family transcriptional regulator; polyprenyl synthetase; secreted beta-mannosidase.
 SOURCE Streptomyces coelicolor A3(2).
 ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 14922)
 AUTHORS Redenbach, M., Krieser, H.M., Denaparte, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.
 TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Mol. Microbiol. 21 (1), 77-96 (1996)
2 (bases 1 to 14922)
Seeger, K.J. and Harris, D.
Unpublished
3 (bases 1 to 14922)
Parkhill, J., Barrett, B.G. and Rajandream, M.A.
Direct Submission
Submitted (04-OCT-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>)
CDS are numbered using the following system eg SC787.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nh1.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid F73 overlaps with cosmid F11 on the AseI-F genomic
restriction fragment.

FEATURES

Location/Qualifiers
1. 14922
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid F73"
1. 122
/note="nominal overlap with StfII"
complement(1..983)
/gene="mana"
complement(1..983)
/gene="mana"
complement(1..983)
/note="SCF73.01c, mana, probable secreted
beta-mannosidase, len: 327 aa; similar to many e.g.
MANA, STBL1 P51529 mannan endo-1,4-beta-mannosidase
precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189
z-score: 1332.7 E(1): 0, 61.1% identity in 288 aa overlap.
Contains N-terminal signal sequence, and Pfam match to
entry PF00150 cellulase, Cellulase (glycosyl hydrolase
family 5). Also similar to SC2H4.16 (62.2% identity in 288
aa overlap)."
/codon_start=1
/transl_table=11
/product="putative secreted beta-mannosidase"
/protein_id="CAB57406.1"
/db_xref="GI:6013077"
/translation="MRPARDPRTLPRLTGLALIGLVVVGALCPGALAAQSPA

misc_feature
gene
CDS

RBS
gene
CDS

RSAGVPSAGQAAGLHIGDRLLGNGDNFVARGVNHATHTYPGETOSLADYKALCAN
SVRVYLDGHRNSENPGADVAAYVTGCKANRLICYLEVDHTDGAAGAAGLLDHAAD
YVIGLADYLAGGEDYIYVINGNPNRNTDPACHTDEPTVAAYKKRAAGQRIIMDAP
NMGQDQVYARAKASVYDADPTGNLISFIMHTSVYDIOEYTDYINAAVDELPLLI
GEGFGPADYGDPEDTMAATAEQLRLGYLAWSWSGNTDPVLDALDFPS"
complement(989..992)
complement(1129..2274)
/gene="SCF73.02c"
complement(1129..2724)
/note="SCF73.02c, possible membrane-bound oxidoreductase,
len: 531 aa; some similarity to eukaryotic
molybdenum-containing oxidoreductases e.g. SMOX RAT Q07116
sulfitc oxidase precursor (488 aa), fasta scores: opt: 262
z-score: 281.9 E(1): 2.5e-08, 30.4% identity in 319 aa
overlap, and NADH-NADPH nitrate reductase 1 (916 aa),
fasta scores: opt: 247 z-score: 262.1 E(1): 3.1e-07, 28.9%
identity in 336 aa overlap. Contains several membrane
spanning hydrophobic domains near N-terminus, and Pfam
match to entry PF00174 oxidored_molpb, Oxidoreductase
molybdoprotein binding domain"
/codon_start=1
/transl_table=11
/product="putative membrane-bound oxidoreductase"
/protein_id="CAB57407.1"
/db_xref="GI:6013078"
/translation="MSDHKTRNRSARFRTWLRGQALSGLLAGCAALAVETLRN"
PNSGPVAVAGGAIIDRTPTAVADMLRFTGIDKLVLGIIAVLLEFLAAGNATV
HRRAGAGLVGIVGAAALGRPDSTGTVGAPVAGVAVLGLAGLLEPFL
EABERGECDRRRFFVIAATAAASAGTGVGRLAGAGREAVASRENTLPDS
RAPVAPROARVEGSSPTPTSDPRYDVALVYKVALATRLRIHGEVAREBTI
TDDRLRLRLDITLTCVSNVSGFPYGNARNICVRLADLACGVPPAGGAD
QVARSVDGMTIGSPVEDWDRDRLAAGNGEPLPEDFGFPVRVAVGLGIVSAC
KMIIELEITPDSYDAVYKRWARPAPYKQSRIDTPPRPAPAGIVAGVAVW
HGEIDKEVVRVDGWEAEVLAEDSRDTWRMSYAMRATKGGHLLTVRALDRASVW
TKRRTVVDGASGRHSVYTVD"
complement(2777..4300)
/gene="SCF73.03c"
complement(2777..4300)
/note="SCF73.03c, unknown, len: 507 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.03c"
/protein_id="CAB57408.1"
/db_xref="GI:6013079"
/translation="MTAFILVSGMFTGTHIMQDTAARLTARGAEVHTVALTGIDAPRA
AAAPGVDETHIADYLAIVDSVGAAGDRITVVGHDYGIHPAVGAADRAREVDRIH
IDSGPDRGVPLAAVPPDSSLERLAGAAGAGVYPPRAHEPFRMGSTAGVPDADLD
RLTALAAPPLDGLLQPLRLTGAVDPVPTGVLTGNGTSIELVQMLVRLGDPALPL
TDPVRSFEFLPTGHPMMLSLPAELTDVYLIRAGAGRLLEPDDTAGHPLFLMDY
PDVPRERHGNLDLYLPDAEPPAVLVVAGGVPADARTPPDMGGLGTCYARCVAGDG
AVGALLDHLHDLDGFERAAADVAAVEVRADPRVDGVRVLAFFSGGGLTAAADLD
APPAMRLCIAATYVPIALPLPNWGSSESRPRAVANGALGVLTTRVGRMPLEAT
VEEFLAAKDCADVVDVDPGHGHGFEITIDTDSRTAVRAATVLDHAGGAR"
complement(4311..4315)
4382..4387
4392..4394
/gene="SCF73.04"
4392..5414
/gene="SCF73.04"
/note="SCF73.04, probable transcriptional regulator, len:
340 aa; similar in N-terminus to members of the Merr
family e.g. NOXA-BRASN nodulation protein NOXA (237 aa),
fasta scores: opt: 211 z-score: 233.9 E(1): 1.2e-05, 48.5%
identity in 68 aa overlap. Also similar to SCD17.06c
(34.8% identity in 322 aa overlap). Contains Pfam match to
entry PF00376 merr, Bacterial regulatory proteins, merr
family"
/codon_start=1
/transl_table=11
/product="putative Merr-family transcriptional regulator"
/protein_id="CAB57409.1"

gene
CDS

RBS
RBS
gene
CDS

/translation="MRPARDPRTLPRLTGLALIGLVVVGALCPGALAAQSPA

gene

/db_xref="GI:6013080"
/translation="MTNSDGLCGTIGELAEAGVTVKTYREYSDRGLLPEASRSAGSH
RYRGEPALDRRLIRLSIALSLGCPVEVRRIIDERHAGTRAGEEGALEAVACGLREV
GESLLAKRKREALRVLEAPPPQADQLRTLGVAVSPPSTVLVRRWGLPRFMRRA
KSAGFLDLVVAVQPDPDEPAQAQYLAFALNLTLPACPVTGPQQPAAHRARSGAAGA
ILYTILAAYEIVLAGVMGRGERPHREGMLDAIFVASYSAYSVCVRTPGFRRLAGOLAAA
PLRDLYGEIYAEEVTTTAGRAPREPFGSAHDMLLAALTAGAVPDGTSPROSKRAY
SPLGV"

/complement(5357..5623)
/gene="SCF73.05c"
/complement(5357..5623)
/gene="SCF73.05c"
/note="SCF73.05c, unknown, len: 88 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.05c"
/protein_id="CAB57410.1"
/db_xref="GI:6013081"
/translation="MDEVLFMKRGVRYGADHNPCBPARGVYAHLVGGPRLDLLVDITD
LTEOERGAVTLATEIGRYGPGGSATYPPEGDTREFPMWRDVP"
/complement(5696..6697)
/gene="SCF73.06c"
/complement(5696..6697)
/gene="SCF73.06c"
/note="SCF73.06c, unknown, len: 333 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.06c"
/protein_id="CAB57411.1"
/db_xref="GI:6013082"
/translation="WLNDIPVTVDGRKRDIPTYWEERPAMTEIAICWTLEARLAHGIR
PVGAARKVRILRPDSVILLRFUCVAGDDAWVKASPAPGAEFAALLTAISLRVPSHYIEPIA
LDNAKKSWLLPDGGPLLFDVIAREVDGDMEBELROYANAOHALTSASEIERLAYEA
APAQAQGVEDLTRDALTPESARDRALSQLBPRILDWCSEILTALGPSDSLHAJDHGGG
OLEFRAAGFTFEWDGADVVOPTGFSSLVPPARAERYGPQVLIPLRNAYTIPEPTWGSS
PRIAQLRRAVSIAMLRLSNLRPAAYAGRILEPGAASHCAATAAGARCILLEDERPL"
/complement(6778..9000)

CDS

gene

gene

gene

Db	397	CGACCGTCGCGCGCGTCAGAGACGTCGCGCGCGCGCGCTCCAGACACGATCATGCTGG	338
Qy	482	atgctgaggagtgaggacaaatttcaacaatcg---atcatgattatggaaggagttt	538
Db	337	ACGCGCCCAACTGGGGGCGAGACGTGAGAGGGCGGTCAATGCGGTGCCACGACGCGTCT	278
Qy	539	ttaatgctgaccctcaagaatacaatgcttttcgattcatatgta	584
Db	277	ACGAGCGCGACCCACCGGCGACCTGATCTTCTCATCCACATGTA	232
RESULT_13			
LOCUS	166494	7218 bp	DNA
DEFINITION	Sequence 14 from patent US 5670367.		PAT
ACCESSION	166494		
VERSION	166494.1	GI:2724471	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 7218)		
AUTHORS	Donner, F., Schellfänger, F. and Falkner, F. Gunter.		
TITLE	Recombinant fowlpox virus		
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;		
FEATURES	Location/Qualifiers		
source	1..7218		
BASE COUNT	1944 a 1491 c 1486 g	1929 t	368 others
ORIGIN	/organism="unknown"		

```

Query Match 3.8%; Score 57.4; DB 2, Length 14922;
Best Local Similarity 46.9%; Pred. No. 0.024;
Matches 273; Conservative 0; Mismatches 296; Indels 15; Gaps 3.

QY 11 ccgatttattatgaagcgggtacacactctataagatgcaatgaaacccatttgaatga 70
Db 814 CCGGCTGCACATCGGTGATCGCGGCTGCTGGAGGGCAAGGCAACGACTTCGTATTC 755

QY 71 gaggatataaccatgycacgcacatggtataaagaccaggaactactgcatttgaagga 130
Db 754 GCGGGGTCAACCAACCCACACACTGGTATCCCGGCAGACGAGTCGCGGGAGGTCA 695

QY 131 ttgcataataccggtgtctaataagrcgggattgtgttactgtatgtggtggaatgtgaca 190
Db 694 AGGC--GCTGGGGCGAAGACAGCTCCGGGTGCTCTTCACAGGGGACCCGCTGGAGG 638

QY 191 aagatgacatccataccagtaagaacactatctctttagcgggaagaatacatattggtg 250
Db 637 AGAAGGCCCGCGGCGACGTCGCCGCGTCATCGAGACATGCAAGGCCAACCGGCTCATCT 578

QY 251 ctgttcctgaagtcaatgatgtctacgcgttatgattcattgc-----ttgccta 301
Db 577 GCGTCTGGAGGTGACGACGACACACCGGCTAAGCCGAGGACCCCGCGGCGACGCTTGG 518

QY 302 atcgcgtcgttgattatttgatttgaataatgagaagtctttaaattggaagaagatacgc 361
Db 517 ACCACGCGCGCCGACTACTGTGATCGGCTCTAAGGACGTCCTGGCGCGGCCACAGGAGATTACG 458

QY 362 tcatattataattgcgaatgaatggttgtgtcgtggaaggggatagtctgggtcagc 421
Db 457 TCATGTCAACATCGGCACAGACCGCTGGGGCAACACGACACCCCGCGGCTGGACCGAGC 398

QY 422 ggtataacaagcaatccgcgattgcgttaacgcggtctaaacacataccttgatgtg 481

```

[illegible]

```

SOURCE          human.
ORGANISM        Homo sapiens
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE           1 (bases 1 to 111071)
JOURNAL         DOE Joint Genome Institute.
REFERENCE       Sequencing of Human Chromosome 16
AUTHORS         Unpublished
TITLE           2 (bases 1 to 111071)
JOURNAL         DOE Joint Genome Institute.
REFERENCE       Direct Submission
AUTHORS         Submitted (29-SEP-2000) Production Sequencing Facility, DOE Joint
TITLE           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT         On Jul 18, 2000 this sequence version replaced gi:7705016.

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----

Project Information
Center Project Name: 602754
Center clone name: RPCI-11_511G21
-----

Summary Statistics
Consensus quality: 104021 bases at least Q40
Consensus quality: 108348 bases at least Q30
Consensus quality: 109319 bases at least Q20
Estimated insert size: 118930; agarose-ef estimation
Estimated insert size: 110521; sum-of-contigs estimation
Quality coverage: 8.44 in Q20 bases; agarose-ef estimation
Quality coverage: 9.08 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 3999: contig of 3999 bp in length
* 4000 4099: gap of unknown length
* 4100 6224: contig of 2125 bp in length
* 6225 6324: gap of unknown length
* 6325 14602: contig of 8278 bp in length
* 14603 14702: gap of unknown length
* 14703 54245: contig of 39543 bp in length
* 54246 54345: gap of unknown length
* 54346 58569: contig of 4224 bp in length
* 58570 58679: gap of unknown length
* 58680 59480: contig of 811 bp in length
* 59481 59583: gap of unknown length
* 59584 60683: contig of 1103 bp in length
* 60684 60783: gap of unknown length
* 60784 62950: contig of 2167 bp in length
* 62951 63050: gap of unknown length
* 63051 64092: contig of 1042 bp in length
* 64093 64192: gap of unknown length
* 64193 85067: contig of 20875 bp in length
* 85068 85167: gap of unknown length
* 85168 91149: contig of 5962 bp in length
* 91150 91236: gap of unknown length
* 91237 108236: contig of 16987 bp in length
* 108237 11071: gap of unknown length
* 10837 11071: contig of 2735 bp in length.
location/Qualifiers
    1..111071
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="16"
        /clone_id="RP11-511G21"
        /clone_1lb="RPCI human BAC library 11"
        /size 27517 c 28256 g 28944 t 1202 others

```

	ORIGIN
Query Match	3.3%; Score 45.8; DB 53; Length 111071;
Best Local Similarity	34.7%; Pred.No. 0.91;
Matches 110; Conservative	0; Mismatches 207; Indels 0; Gaps 0;
OY 232 gaagataatcatttggtgctgcttctaagtcaatgcataacccggatgatgccatt 291	
Dd 59292 GATGGTGATGAGTGTGAATAATATGATGATGATGATGCACAGGGGATGATGATGCCAT 59351	
OY 292 gcttcgcacaactcgctgccttatgattcatgtaagaatggaaggctttaatgaaag 351	
Dd 59352 GATTGAATGTAATGCTGATGCGTCATGATGCCCATGATGATGTAATGCTTTGCTATGCT 59411	
OY 352 gaagataaccgtcataattaataatitcgaatatgatgtttgctcgtggaaaggagatgct 411	
Dd 59412 GATGGTGATGCTGATGAGTAATGTAATCTCATGCTGATGATGATGCTGCTGATGCTATGCT 59471	
OY 412 tgggctgacgggtataaacaagcaatcccgcgattgctglaaccgctctaacataacc 471	
Dd 59472 GATGCTGATNNN 59531	
OY 472 ttgtagtgtatagatgcctgcggggtgggacaaattccaacatgatcatcatatgaaaga 531	
Dd 59532 NNN 59591	
OY 532 gaagattttaatgctga 548	
Dd 59592 GATGCTGATGATGATCA 59608	
RESULT 15 CELK03E6 LOCUS CELK03E6 35049 bp DNA INV 28-OCT-1998 DEFINITION Caenorhabditis elegans cosmid K03E6. ACCESSION U55375 VERSION U55375.1 GI:3805671 KEYWORDS SOURCE ORGANISM . Caenorhabditis elegans. Caenorhabditis elegans. Euharyota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea; Rhabditiidae; Pelodertinae; Caenorhabditis. 1 (bases 1 to 35049) Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bofield,J., Burton,J., Connell,M., Cosey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,I., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurtry,A., Mortimore,B., O'Callaghan,P., Parsons,J., Percy,C., Ricklen,A., Roopra,A., Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonhammer,E., Staden,R., Stenson,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Spoat,J. and Wohlman,P.	
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans .
JOURNAL MEDLINE	Nature 368 (6466), 32-38 (1994)
REFERENCE	94150718.
AUTHORS	2 (bases 1 to 35049) Latreille,P. and Gatlung,S.
TITLE	The sequence of C. elegans cosmid K03E6
JOURNAL	Unpublished (1998)
REFERENCE	3 (bases 1 to 35049)
AUTHORS	Waterston,R.
TITLE	Direct Submission
JOURNAL	Submitted (17-APR-1996) Robert Waterston
REFERENCE	4 (bases 1 to 35049)
AUTHORS	Waterston,R.
TITLE	Direct Submission
JOURNAL	Submitted (21-APR-1997)
REFERENCE	5 (bases 1 to 35049)
AUTHORS	Waterston,R.

TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Oct 28, 1998 this sequence version replaced g1:1280123.
Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: r.wenham@code.wustl.edu and j.sanger@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is C04D1, 2700 bp overlap; 3' clone is F49E7, 200 bp overlap. Actual start of this clone is at base position 701 of CELK03E6; actual end is at 35049 of CELK03E6

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES
Location/Qualifiers

1..35049

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/clone="K03E6"

/chromosome="X"

/complement(3691..7748)

/gene="unc-1"

/complement(join(3691..3766,3931..4145,4822..5095,5154..5278,6432..6545,7296..7326,7735..7748))

/gene="unc-1"

/note="K03E6.5"

/codon_start=1

/product="Erythrocyte band 7 integral membrane protein"

/protein_id="AAC69044.1"

/db_xref="gi:1280125"

/translation="MEYSHQIQPKSCATIDVPDYETIGTIFGVALQALMILLIVTFPMSVCIVKIVKERYVIFRIGLVEFGAGSGMIFIIPIIDYRKIDLVAVP

POELSKDSVAVDAVYFRTSDIAGVNVDAIYFTKLLAQDTLNLGAMTITE

MLTEREATACICETLIDGTEHMGYKVRVVKRIKRIPOQTRMAAARARAK

VVAEGEKASRALEADAVIQANFVALQHLNLAHNSIAHNSITVFVPEFEGA

FMKRDQ"

/complement(12649..14893)

/gene="K03E6.4"

/complement(join(12649..12750,12801..12928,13397..13527,13570..13666,14217..14391,14837..14893))

/gene="K03E6.4"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC69043.1"

/db_xref="gi:1280126"

/translation="MEKADSVITLHGVRTPTEPESNVCKLSTVATGDKKLINGE

TVSLSSEFRGADNIFIDLPFGCLKSSSPFMPRIPLANKFSKDHSGKDCVYTGML

PIFTEPGHVCVRHCFEPSPSTVQQLAKLRVNAEVPMLPANAAGACRSTIK

GYDLIEINVSICSDTAKYKSTSTSEIKSSNSHANNSTIADLAGIEAVGNETFN

AQNDNEDESKR"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

gene 20796..24499
/gene="11m-6"

Db 33970 CTTTATAGGTCATGCCAATTTTGTGTGTGTTTGTCTGGAAGTAAACAATTTTAT 34029

Oy 344 ttggaag 351

|||||

Db 34030 GCGGAATG 34037

Search completed: December 20, 2000, 03:20:08
Job time: 13253 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:58:48 ; Search time 168.09 Seconds
(Without alignments)
3084.148 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_1470
Perfect score: 1380
Sequence: 1 aatcaaatccgattttt.....ttgataatgatgttagaa 1380

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /\$XDS6/gcgdata/geneseq/NA1980.DAT:*
2: /SID6/gcgdata/geneseq/NA1981.DAT:*
3: /SID6/gcgdata/geneseq/NA1982.DAT:*
4: /SID6/gcgdata/geneseq/NA1983.DAT:*
5: /SID6/gcgdata/geneseq/NA1984.DAT:*
6: /SID6/gcgdata/geneseq/NA1985.DAT:*
7: /SID6/gcgdata/geneseq/NA1986.DAT:*
8: /SID6/gcgdata/geneseq/NA1987.DAT:*
9: /SID6/gcgdata/geneseq/NA1988.DAT:*
10: /SID6/gcgdata/geneseq/NA1989.DAT:*
11: /SID6/gcgdata/geneseq/NA1990.DAT:*
12: /SID6/gcgdata/geneseq/NA1991.DAT:*
13: /SID6/gcgdata/geneseq/NA1992.DAT:*
14: /SID6/gcgdata/geneseq/NA1993.DAT:*
15: /SID6/gcgdata/geneseq/NA1994.DAT:*
16: /SID6/gcgdata/geneseq/NA1995.DAT:*
17: /SID6/gcgdata/geneseq/NA1996.DAT:*
18: /SID6/gcgdata/geneseq/NA1997.DAT:*
19: /SID6/gcgdata/geneseq/NA1998.DAT:*
20: /SID6/gcgdata/geneseq/NA1999.DAT:*
21: /SID6/gcgdata/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1380	100.0	1470	21	245335
2	897.2	65.0	1438	21	245336
3	639.4	46.3	1482	21	229845
4	639.4	46.3	1482	21	245337
5	601	43.6	1407	21	229846
6	601	43.6	1407	21	245338
7	408.4	29.6	1107	21	245342
8	405	29.3	995	21	245340
9	304.4	22.1	960	21	245345
10	167.8	12.2	564	21	245346
11	136	9.1	915	21	245343
12	109.4	7.9	1461	10	N91621

13	104.4	7.6	397	21	245344
14	41.8	3.0	2635	18	T83966
15	41.8	3.0	4826	18	V74586
16	40.4	2.9	13321	18	V74315
17	35.4	2.6	6741	21	A10595
18	35	2.5	2040	17	T29774
19	35	2.5	2100	10	N90712
20	35	2.5	2901	12	Q14810
21	35	2.5	3558	13	Q22596
22	35	2.5	9542	20	X20260
23	34.6	2.5	580073	18	T58840
24	33.6	2.4	627	20	Z08233
25	33.6	2.4	1395	20	Z08232
26	33.6	2.4	1680	20	Z08234
27	33.6	2.4	1763	15	Q55316
28	33.4	2.4	2408	19	V17621
29	33.4	2.4	2802	12	Q14809
30	33.2	2.4	633	18	V75647
31	33	2.4	782	17	T10931
32	33	2.4	1289	20	X26810
33	33	2.4	1800	20	Z25012
34	33	2.4	2703	21	A09056
35	32.8	2.4	6247	20	X86997
36	32.6	2.4	141589	21	A34791
37	32.6	2.4	141589	21	A35005
38	32.6	2.4	141589	21	A35030
39	32.4	2.3	1300	20	V82078
40	32.4	2.3	2132	20	V82077
41	32.4	2.3	2841	18	V74488
42	32.4	2.3	3006	21	Z29484
43	32.4	2.3	22080	20	X06751
44	32.2	2.3	1021	16	T04613
45	32.2	2.3	1026	17	T13340

ALIGNMENTS

DNA encoding a ...
DNA encoding a ...
Staphylococcus aur ...
Staphylococcus aur ...
Gene encoding a su ...
Bacillus thuringie ...
cryp gene ...
Btm pg14 72kDa Cry ...
Coding region of ...
Borrelia burgdorfer ...
Mycoplasma genital ...
Lactobacillus brev ...
Coding sequence of ...
Lactobacillus brev ...
Lactobacillus brev ...
Bacillus sp. V230 ...
Btm pg14 72kDa Cry ...
Staphylococcus aur ...
Borrelia burgdorfer ...
orf2 gene sequence ...
Lactobacillus reut ...
Human Th2-specific ...
A. thaliana hydrop ...
Human adenosine re ...
Human adenosine re ...
DNA encoding a par ...
DNA encoding a gid ...
Staphylococcus aur ...
DNA encoding human ...
Salmonella enteric ...
5' flanking region ...
Listeria phage lys

RESULT 1	
ID 245335	245335 standard; DNA; 1470 BP.
XX	
AC 245335;	
XX	
DT 27-MAR-2000	(first entry)
XX	
DE	DNA encoding a Bacillus mannanase enzyme.
XX	
KW	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW	endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW	galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW	printing paste; plant material degradation; recycled waste paper;
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;
KW	mannan-containing food; coffee extract; cleaning composition;
KW	machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW	contact lens; body-care composition; fabric softener; oil well drilling;
KW	subterranean formation fracture; ss.
XX	
OS	Bacillus sp.
XX	
FH	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1470
FT	/tag- a
FT	/product- "mannanase"
FT	/trans_except- (pos: 346..348, aa: Pro)
FT	/note- "no termination codon given"
FT	1..90
FT	/tag- b
FT	91..1470
FT	/tag- c
FT	/note- "specifically claimed in claim 4; nucleotides 91-990 encode the catalytically active core"

|||||
Db 1351 tctcagattcacaagtcacaaatctctcctcaagcaagggaattcggatctcagttc 1410
Qy 1321 caatcagagtgatgagtagtgagacaacatcagattatattgataatgattgtagaa 1380
Db 1411 caatcagagtgatgagtagtgagacaacatcagattatattgataatgattgtagaa 1470
RESULT 2
245336
ID 245336 standard; DNA; 1438 BP.
XX 245336.
AC
DT 27-MAR-2000 (first entry)
XX
DE DNA encoding a mannanase-linker-cellulose binding domain fusion.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; Bacillus sp. 163; galactomannan;
KW 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW printing paste; plant material degradation; recycled waste paper;
KW paper making pulp; guar; locust bean gum; thickener; viscosity;
KW mannan-containing food; coffee extract; cleaning composition;
KW machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW contact lens; body-care composition; fabric softener; oil well drilling;
KW subterranean formation fracture; cellulose binding domain; ss.
XX
OS Synthetic.
OS Bacillus sp.
OS Clostridium thermocellum.
XX
FH Key Location/Qualifiers
FT CDS 1..1431
FT /tag- a
XX
PN W09964619-A2.
XX
PD 16-DEC-1999.
PE 10-JUN-1999; 99W0-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Kauppinen MS, Schneelein M, Schmorl K, Andersen LN, Bjornvad ME;
XX
DR WPI; 2000-105891/09.
DR P-PSDB; Y54123.
XX
PT New mannanases for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions
XX
PS Example 4; Page 210; 242pp; English.
XX
CC The present sequence encodes a mannanase-linker-cellulose binding
CC domain fusion protein. Mannanase (also known as mannan
CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)

CC hydrolyses galactomannans. Specifically, mannanases hydrolyse
CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
CC and galactoglucomannans. The mannanase protein, or preparations
CC containing it, are used to improve properties of cellulosic or
CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sites or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC celluloses; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.
XX
SQ Sequence 1438 BP; 445 A; 267 C; 344 G; 382 T; 0 other;
Query Match 65.0%; Score 897.2; DB 21; Length 1438;
Best Local Similarity 99.7%; Pred. No. 6,2e-248;
Matches 899; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 4 gcaattcggatttattgtaagcgtacacactctatcgatgcgaatggaacccattt 63
Db 1 gcaattcggatttattgtaagcgtacacactctatcgatgcgaatggaacccattt 60
Qy 64 gtaatgagaggaattaaacattggtgacacatggtataagaccggacacactgcaatt 123
Db 61 gtaatgagaggaattaaacattggtgacacatggtataagaccggacacactgcaatt 120
Qy 124 gaaggagattgcaaataccggtgctaatacgggtccggattggtattcgtatgaggagaca 183
Db 121 gaaggagattgcaaataccggtgctaatacgggtccggattggtattcgtatgaggagaca 180
Qy 184 tggaaaaagatgacatccatcagtaagaaccttactctttagcggagaataatcat 243
Db 181 tggaaaaagatgacatccatcagtaagaaccttactctttagcggagaataatcat 240
Qy 244 ttggttgcgtcttctgtaagtcattgatcccggttatgattccatgcttcgctcat 305
Db 241 ttggttgcgtcttctgtaagtcattgatcccggttatgattccatgcttcgctcat 300
Qy 304 cgtgctgttgatttattgtaattggaatggaagtgtcttaattggaaggaatacgtc 363
Db 301 cgtgctgttgatttattgtaattggaatggaagtgtcttaattggaaggaatacgtc 360
Qy 364 attattaatttgcgaatggaatggttgcgtcgtggaagggatggttgcgtgacgg 423
Db 361 attattaatttgcgaatggaatggttgcgtcgtggaagggatggttgcgtgacgg 420
Qy 424 tataaacaagcaatcccggtattggtiaacccggttcaaacacttcatgattagat 483
Db 421 tataaacaagcaatcccggtattggtiaacccggttcaaacacttcatgattagat 480
Qy 484 gctgcgggttggtgacaattccacaatcgaattcagttatggaagaagaagttttaat 543
Db 481 gctgcgggttggtgacaattccacaatcgaattcagttatggaagaagaagttttaat 540
Qy 544 gctgacctcaacgaataacatgtttcgaattcatatgatatgacaggtgtat 603
Db 541 gctgacctcaacgaataacatgtttcgaattcatatgatatgacaggtgtat 600
Qy 604 gcatcgcaagttcgtactcaatattgacgaagttcttaatacgaacctcgattagatt 663
Db 601 gcatcgcaagttcgtactcaatattgacgaagttcttaatacgaacctcgattagatt 660
Qy 664 gttgaatttgacacggtcatcaaatatgttgcgtgataagcaacgattatgacctat 723
Db 661 gttgaatttgacacggtcatcaaatatgttgcgtgataagcaacgattatgacctat 720
Qy 724 tctgaacaagaaggagttggtgtgtgtggtcgtgtaatggaagggaacgcccagaatgt 783

ID	Sequence	Location/Qualifiers
Db	721 tcgcgaacaagaagagctgtggttggttcgctgctgcacgtgaagggaagccagcatgtg	
Qy	784 gagtatttaaacctttcgaatgatgtggcctggaaataaccttaacgtttgggaaataca	843
Db	781 gagtatttaaacctttcgaatgatgtggcctggaaataaccttaacgtttgggaaataca	840
Qy	844 atagtgatgtgctcatatgtgttaagaagaaacttcgagatgaagcaccgtttttacagct	903
Db	841 atagtgatgtgctcatatgtgttttaagaagaaacttcgagatgaagcaccgtttttacagct	900
Qy	904 gg 905	
Db	901 ag 902	
RESULT	3	
ID	229845	
XX	229845 standard; DNA; 1482 BP.	
XX	229845;	
DT	27-MAR-2000 (first entry)	
XX	Bacillus agaradherens NCIMB 40482, Mannanase encoding DNA.	
DE	Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;	
KW	detergent composition; mid-branched anionic surfactant; washing;	
KW	cosmetic stain; food stain; ss.	
XX	Bacillus agaradherens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..1482
FT		/*tag- a
FT		/product- "Mannanase"
FT		/EC_number- "3.2.1.78"
FT	sig_peptide	/function- "Hydrolysis of 1,4-beta-D-mannosidic linkages"
FT		1..96
FT		/*tag- b
FT	mat_peptide	97..1029
FT		/*tag- c
FT		/label- Mature_Mannanase
PN	W09964552-A1.	
XX	16-DEC-1999.	
PD		
XX	10-JUN-1998; 98MO-US12026.	
PE		
XX	10-JUN-1998; 98MO-US12026.	
PR	(PROC) PROCTER & GAMBLE CO.	
XX		
PA	Betitol JP, Thoen CAJK;	
XX		
PI	WPI: 2000-116536/10.	
DR	P-PsDB; Y44495.	
DR		
XX	Detergent composition for removing greasy stains such as cosmetics,	
PT	food stains and body soils	
XX		
PS	Disclosure; Page 101; 113pp; English.	
XX		
CC	The present sequence is a DNA encoding B. agaradherens NCIMB 40482,	
CC	alkaline mannanase enzyme. Mannanase shows maximum activity at pH ranging	
CC	from 7.5-10.5. It can be used in a detergent composition along with a	
CC	mid-branched anionic surfactant. The detergent composition may be used	
CC	for washing purposes, soaking/pre-treatment of stained fabric, hard	
CC	surface cleaning and for removal of cosmetic and/or food stains. This	
CC	composition provides excellent cleaning effect at low temperature.	
SO	Sequence 1482 BP; 446 A; 285 C; 352 G; 399 T; 0 other;	

Query Match	SLimilarity	46.3%	Score	639.4	DB	21	Length	1482	
Best Local	SLimilarity	67.5%	Pred.	No. 6.7e-117					
Matches	933	Conservative	0	Mismatches	441	Indels	9	Gaps	2
QY	4	gcaaatccggaatttattgtaagcggtaaccatactacatacgaatgccaatggaaccat	63						
DB	94	gaaagtacagcgtttattgtattgtatgacatacgttatatgacgaatggaatggacgat	153						
QY	64	gtaatgagagagatttaaccatatgagcagcatatgtataagaccaggaactcttcgat	123						
DB	154	gtcaatgagagattattaaacctatgacaatcttctgtataaagaaacccgttcaacat	213						
QY	124	gaagagattgccaataccggtgtcatatcggtccgagattgttatactgattgggagaca	183						
DB	214	ccctgcatttgcagagcaagggcgcacaacacgatctgtattgtttatccgatgvcgtca	273						
QY	184	tggacaacaaagatgacatccatacagtaagaacacttactctttagcggaaatacat	243						
DB	274	tgggaaaaaagaaagacatgacacacctctgtgaagtcgatattgagcttgcgagacaataa	333						
QY	244	ttagtgcgtcttcttgaagttcatactatgcttaccgggttatgattccattgcctgcctcat	303						
DB	334	atgvgcgtctgccttgcgaagttccatgacatgcacaggggtccgcgattccgcatat	393						
QY	304	cgctgcctgtatatttggatttgaatgtgaaagtgttctaatttgaagaagaaatccgtc	363						
DB	394	cgaagcgtgtattatttgaatagaagaagtgccttactcgttaagaagaatcaggt	453						
QY	364	attataatattgcgaatgaagatggtttgtcttcgtggaaaggagatgcttggttcaggg	423						
DB	454	attataatcatttgcacaacagatgctgatagtggatgtggatgcttcagcttggccgagtc	513						
QY	424	tataaacaagacatcccgatctgctgaacgcgggtctcaaaacataccttgaattgtaag	483						
DB	514	tatatgtatgcatcttcggaagcttcgcgatgctgcggcttaacacacacttaattgttgat	573						
QY	484	gtctcgagggtggggacaattccacaatcagatcatgatatatgagaagagattttaat	543						
DB	574	gcaagagagatggggacaatccgcacatctatcttcatgattacggacaagaatgtgtta	633						
QY	544	gtgcgccttcacgacaataaatggttttcgatcttcaatgatagaataatgacaagtgtgat	603						
DB	634	gcaagtcctgtaaaaaataacagatgtcttccatccataatgatatgacgtgcgtgtgat	693						
QY	604	gcatcgcgaagtctgtactaataatgtgaccagagttctttaaacaagaccctgcatactcat	663						
DB	694	gtaaacactgttagatcataatattgatatagatgcatatagaacacactcgtctctgtaata	753						
QY	664	ggtgaatttggacaacgcgcataaacaattgtgacgtcgttgaagcaacgatatatgagcat	723						
DB	754	ggtgaatttcggcacaagaacatactgataatgtgattgttgaagatacaatcccttgatcat	813						
QY	724	tctgacaacaagagaggttgggtgtgttggcgtggtcatgtgaagagggagagcccaagtg	783						
DB	814	tctgagaagaacctgacagaggtgcgtcgtctgttcttgtaaagagacaagaatccgaatgg	873						
QY	784	gaagtatttaagaccttcgataatgataatgggctcggaaataaacccttaacagcttgggaaataca	843						
DB	874	gaattatttaagaccttccagaagaactcgtggtctgttcaacaacttaacatgatttggggagataga	933						
QY	844	atagagaattgtccataatggttttaagagaacaacttcagagattbaagacaacgtttttacaagt	903						
DB	934	attgtccacgcgggcccgaatggcttaagaagaaacctccaacacatccacacgttatcttcagat	993						
QY	904	-----ggaggaatctgatatgagagaaactctcccgacaacatcttataatttgaagtgtg	957						
DB	994	gataacggtgtgcacaccctgcgaacgcgaacactgtactacttgaatttgaaggaagc	1053						
QY	958	atgcaagagatggaactcgaagaatgacttgagcggagagttccttggctgtgacagagatgtctc	1017						
DB	1054	acacaagaggtgtgcatacgaagaacagcttgacccgtgtgcccctgtctccgttaacaagaatggggt	1113						

```

PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99DS-0123543.
PR 10-MAR-1999; 99DS-0123623.
PR 10-MAR-1999; 99DS-0123641.
PR 11-MAR-1999; 99DS-0123642.
XX
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX
XX Kauplien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX WPI: 2000-105891/09.
XX P-PSDB: Y54124.
XX
XX New mannases for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions.
XX
XX Example 5; Page 212-213; 242pp; English.
XX
XX
XX The present sequence encodes a Bacillus mannanase (also known as mannan
XX endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The
XX mannanase hydrolyses galactomannans. Specifically, mannases hydrolyse
XX 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
XX and galactoglucomannans. The mannanase protein, or preparations
XX containing it, are used to improve properties of cellulosic or
XX synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannases
XX are also used to process coffee extracts (to inhibit gel formation);
XX in cleaning compositions (for machine washing of fabrics, as
XX hard-surface cleaners, for hand or machine dishwashing, also in oral,
XX dental, contact lens or body-care compositions) where they remove
XX mannan-containing soils and prevent binding of some soils to
XX celluloses, and in fabric softeners. They can also be used in oil
XX well drilling to fracture subterranean formations.
XX
XX Sequence 1482 BP; 446 A; 285 C; 352 G; 399 T; 0 other;
SQ

```

```

QY 424 tataacaagcaatccgcgattggtacgcccgtctaaacattacttgtagat 483
DB 514 tatattgtagcattccgaagcttcgcgattccggtctaaacacacattactgtagat 573
QY 484 gctgcggggtgggaacattccacaacatcgattatgattatggagaagattttaat 543
DB 574 gcagagagatggggcaatcccaatctatctatgattacggaagaagatggtttaat 633
QY 544 gctgacccccaagcaatatactgtttcgtattcattgattgattatgacgggtgaat 603
DB 634 gcagatccgttaaaaaatacgaattctccatccatagatgattgctggtgtagat 693
QY 604 gcatcgcaagttcgtaactaatatgacaggttcttaacagaccctcgatgtagat 663
DB 694 gctaacaactgtagatcaaatattgtagatgtagatgtagatgtagatgtagat 753
QY 664 ggttaatttgtagacccgtcattacaaatgtagatgtagatgtagatgtagat 723
DB 754 ggtgaattcggtcattacaaatgtagatgtagatgtagatgtagatgtagat 813
QY 724 tctgaacaagagagatgtagatgtagatgtagatgtagatgtagatgtagat 783
DB 814 tctgaagaagaactgtagatgtagatgtagatgtagatgtagatgtagatgtagat 873
QY 784 gactatttagaacttctgaatgtagatgtagatgtagatgtagatgtagatgtagat 843
DB 874 gactatttagaacttctgaatgtagatgtagatgtagatgtagatgtagatgtagat 933
QY 844 atagtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 903
DB 934 atgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 993
QY 904 -----ggagagatcgttagatgtagatgtagatgtagatgtagatgtagatgtagat 957
DB 994 gataacggtgtagatcctgtagatcctgtagatcctgtagatcctgtagatcctgtagat 1053
QY 958 atgcaagagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1017
DB 1054 acacaagaggtgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1113
QY 1018 tctaaagagatccttctaaagcgtagatccttctaaagcgtagatccttctaaagcgtagat 1077
DB 1114 gctcagtagatccttctaaagcgtagatccttctaaagcgtagatccttctaaagcgtagat 1173
QY 1078 tctcagtagatccttctaaagcgtagatccttctaaagcgtagatccttctaaagcgtagat 1137
DB 1174 cgtatagtagatccttctaaagcgtagatccttctaaagcgtagatccttctaaagcgtagat 1233
QY 1138 catgcaaatggggaaggtgtgtgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1197
DB 1234 catgcaaatggggaaggtgtgtgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1293
QY 1198 catggtatatacagtagatcgtgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1254
DB 1294 tctgattatatacagtagatcgtgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1353
QY 1255 aagcgtatcctgtagatccttctaaagcgtagatccttctaaagcgtagatccttctaaagcgtagat 1314
DB 1354 aagcgtatccttctaaagcgtagatccttctaaagcgtagatccttctaaagcgtagatccttctaaagcgtagat 1413
QY 1315 cagttcacaacagtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1374
DB 1414 caatttcagcgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1473
QY 1375 gta 1377
DB 1474 tta 1476

```

RESULT 5
 229846
 ID Z29846 standard; DNA: 1407 BP.
 XX

```

AC 229846;
XX
XX 27-MAR-2000 (first entry)
DE Bacillus agaradherens Clone MB594, Mannanase encoding DNA.
XX
XX Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KW detergent composition; mid-branched anionic surfactant; washing;
XX cosmetic stain; food stain; ss.
XX
OS Bacillus agaradherens.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..1407
XX FT /*tag= a
XX FT /product= "Mannanase"
XX FT /EC_number= "3.2.1.78"
XX FT /function= "Hydrolysis of 1,4-beta-D-mannosidic linkages"
XX FT sig_peptide 1..93
XX FT mat_peptide 94..1404
XX FT /*tag= b
XX FT /*tag= c
XX PN W09964552-A1.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Bettiol JP, Thoen CAJK.
XX
XX WPI: 2000-116536/10.
XX DR P-PSDB; Y44496.
XX
XX Detergent composition for removing greasy stains such as cosmetics,
XX food stains and body soils -
XX
XX Disclosure; Page 102-103; 113pp; English.
XX
XX The present sequence is the B. agaradherens Clone MB594 DNA encoding
XX alkaline mannanase enzyme. It is derived from B. agaradherens strain
XX NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
XX from 7.5-10.5. It can be used in a detergent composition along with a
XX mid-branched anionic surfactant. The detergent composition may be used
XX for washing purposes, soaking/pre-treatment of stained fabric, hard
XX surface cleaning and for removal of cosmetic and/or food stains. This
XX composition provides excellent cleaning effect at low temperature.
XX
XX Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other;
SQ

```

Query Match 43.6%; Score 601; DB 21; Length 1407;
 Best Local Similarity 67.3%; Pred. No. 6; 9e-163;
 Matches 882; Conservative 0; Mismatches 420; Indels 9; Gaps 2;

```

QY 4 gcaattcggattttagaagcggttacacactctatagatgcaatggaacccatt 63
DB 94 gcaagtagaagccttctatgtagatgtagatgtagatgtagatgtagatgtagat 153
QY 64 gtaatgagaggaattaacatgtagatgtagatgtagatgtagatgtagatgtagat 123
DB 154 gtagagaggaattaacatgtagatgtagatgtagatgtagatgtagatgtagat 213
QY 124 gaaggaattgcaataacggtgtagatgtagatgtagatgtagatgtagatgtagat 183
DB 214 cctgcatgtagaagaagcgtagatgtagatgtagatgtagatgtagatgtagatgtagat 273

```

OY	184	tggcaaaaagatgcaaccctacatgaagaacacctatcctcttagcggaataatcat	245
Db	274	tggaaaaaagcgcacattgcaacacattcggaaagcatctgagcttcggyagacaataaa	333
OY	244	tggctcgtctcttggaagttcatgatgcacacggttatgatccattgcttcgcgtcat	303
Db	334	atggctgctgtctggttgaagttcatgatgcacacggttcgcgtatccgcgcgtatattaat	393
OY	304	cgctgcttgatattggaattgaaatgaaagtgctttaattgnaaaggaaataccgctc	363
Db	394	cgagccgttgattatgatagaatagaaagatgcgcttatcgtgaagaagatacgcgtc	453
OY	354	attattaatatcggaatgaatggtcttgcttgaggaaaggagatgcttgggtcgaagg	423
Db	454	attattaacattgcaaaaacgcggtgtgtaaggagattggatgctgcacgttggccgatggc	513
OY	424	tataaacaagaacacccgcgattgctgaaagccgcggtcttaaacatctccttggttgat	483
Db	514	tatatgtatgtcattccgaagcttcggatgagccgggttaacacacacctaattgttgat	573
OY	484	gctgcgggttgaggacaattccacaaatcgaattcatgatattgaaaggagaagttttaat	543
Db	574	gcagcagatgagggcgaatatccgcaatctattcatgatattacggaacagatggtttaat	633
OY	544	gctacccctcaacgaataacatggttttcgattcatatgatataatgcaggtgtgtaat	603
Db	634	gcgaatccggttaaaaaataataagatgtctccatcccatatgatatagtatgctgtg	693
OY	604	gcatacgcaagttcgttaataattgaccgcggtcttaatacgaagacctgcatagtact	663
Db	694	gctacaacgtgttaagatcaaaatattgataagatgcatatgaatcaaaacctgtctcgtlaa	753
OY	664	ggtgaatttgacaacgcttcatacaaaatgtagcgtgcgatgaaagcaacgataatgagcat	723
Db	754	ggtgaattcgtctcaatagacatactatgtagttagttagaagatacaaatcctcttgat	813
OY	724	tcctgaacaaagagagatgtaggtgtgttgcggtgtcatgtgaaagggaacggcccaaatg	783
Db	814	tcctgaataaacctgcaacaggtgtcgtcgtgtcttgtaaaaggcaacagatccgaatg	873
OY	784	gagagattgaagcctttgaaatgttggtgcgtggaataaactatacagcttgaggaaataca	843
Db	874	gacattatgaaccttcaagaagactggtgcgtgcaacatttaactgtatggtgggaatga	933
OY	844	ataatgaaatgtccataatgtgtttaaagaaatctggaattgaagccggttttaacagt	903
Db	934	atgtccacgggggcgagatggtcttaacagaaacctccaacacatccacgtaatttcagat	993
OY	904	-----ggaagatctgattgtagagaaactctccgcacaactcttatgatatttgaagtagt	957
Db	994	gataacggtgtgtaaccttcgaacctgcgaacctgtaactactcttatgatatttgaagaa	1053
OY	958	atgaagaagatggaactggaatgaacttgcagccggaggtcccttggtgcgtgacaaagtgtct	1017
Db	1054	acaacaagggtgtagcatgtgaagcaacgtgtaacccggtggtcccttggtccgtaacaaatggygt	1113
OY	1018	tctaagaagaatctacttttaaaagcgaattacaatttgcttgcaaatccacaacattac	1077
Db	1114	gtctcaggttaactctctttaaaagccgagatgtaaatttaactccaatcttccacatga	1173
OY	1078	ttaacagttaattcaaaatagctcttaacagcaagaatagtagatacaaaagctactgttaa	1137
Db	1174	ctgtatagtgaacaaagctgtaacttaatacgaatactctcagctcaaaagcaacgcttcgc	1233
OY	1138	catgcaaatgggggaagtgttggttatatgaaatgtaactgcgcgctcttatgtgnaaacgga	1197
Db	1234	catgccaattggggaataatcccggttaagcatgaaatgcaagaactttaaagctgnaaaacgggc	1293
OY	1198	catggttatataatgtagtactctggaagcttctgtgcgattgaacggttc---actcgaaca	1254
Db	1294	ctgtattatataatgtagcatagcgttccctttaaaccgtaataatgataagctccaactcagaaca	1355
OY	1255	acgcatatctagatttatataaatgtccaaacacttcttcaagaatgaaggaa	1305

Db	1354	acgctatccttgattcaacacatcgaaatcatcatcgttagggaaa	1404
RESULT	6		
ID	245338	245338 standard; DNA; 1407 BP.	
AC	245338;		
XX			
DT	27-MAR-2000	(first entry)	
XX			
DE	DNA encoding a <i>Bacillus mannanase</i> enzyme.		
XX	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;		
KW	endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;		
KW	mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibr		
KW	synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;		
KW	plant material degradation; recycled waste paper; paper making pulp;		
KW	guar; locust bean gum; mannan-containing food; coffee extract;		
KW	cleaning composition; machine washing; hard-surface cleaner;		
KW	dishwashing; oral; dental; contact lens; body-care composition;		
KW	fabric softener; oil well drilling; subterranean formation fracture; ss		
XX	Synthetic.		
OS	<i>Bacillus agaradhaerens</i> .		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1407	
FT		/tag= a	
FT		/product= "mannanase"	
XX			
PN	MO964619-A2.		
XX			
PD	16-DEC-1999.		
XX			
PF	10-JUN-1999;	99WO-DK00314.	
XX			
PR	10-JUN-1998;	98US-0111256.	
PR	20-OCT-1998;	98DK-0001340.	
PR	20-OCT-1998;	98DK-0001341.	
PR	28-OCT-1998;	98US-0105970.	
PR	28-OCT-1998;	98US-0106054.	
PR	23-DEC-1998;	98DK-0001725.	
PR	05-MAR-1999;	99DK-0000306.	
PR	05-MAR-1999;	99DK-0000307.	
PR	05-MAR-1999;	99DK-0000308.	
PR	05-MAR-1999;	99DK-0000309.	
PR	09-MAR-1999;	99US-0123543.	
PR	10-MAR-1999;	99US-0123623.	
PR	10-MAR-1999;	99US-0123641.	
PR	11-MAR-1999;	99US-0123642.	
XX			
PA	(NOVO) NOVO-NORDISK AS.		
XX			
PI	Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;		
XX			
XX	WPI: 2000-105891/09.		
DR	P-PSDB: Y54125.		
XX			
PT	New mannanases for treatment of textiles, plant material and coffee		
PT	extract, and in cleaning compositions -		
XX			
PS	Example 5; Page 214-215: 242pp: English.		
XX			
CC	The present sequence encodes a <i>Bacillus mannanase</i> (also known as mannan		
CC	endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).		
CC	The present mannanase is a synthetic variant of the mannanase of		
CC	Y54124, in which the C-terminus of the protein was changed due to		
CC	design of a lower PCR primer used for amplification. The mannanase		
CC	hydrolyses galactomannans. Specifically, mannanases hydrolyse		
CC	1,4-beta-D-mannosidic linkages in mannans, galactomannans,		
CC	glucomannans, and galactoglucomannans. The mannanase protein, or		

PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
PA (NOVO) NOVO-NORDISK AS.
PI
PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
XX WPI: 2000-105891/09.
DR P-PSDB; Y54127.
XX
PT New mannanases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions
XX
PS Example 10; Page 219-220; 242pp; English.
XX
CC The present sequence encodes a *Bacillus* mannanase (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannanases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannan, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.
XX
XX Sequence 995 BP; 286 A; 162 C; 256 G; 291 T; 0 other;

		29.3%;	Score 405;	DB 21;	length 995;
		Best Local Similarity 66.1%;	Pred. No. 1-le-106;		
Matches	585;	Conservative	0;	Mismatches 300;	Indels 0; Gaps 0;
<hr/>					
QY	13	ggatttatgtatgaagcgttcacactctatacgcgatgcaaatgaaaccatttgaatgaga	72		
Db	106	gggtttccatgtttaatgttgtacaacattatgtatgcaaatgaaaaccttttgtagaga	165		
<hr/>					
QY	73	gggattaaccatcgvggcagcacgcatggtataagaagccaggcacactctgcaattgaaggatt	132		
Db	166	gggatttaaccatgagaaatcgttggttaaacaagaactagaaaacatccatgagaaggatt	225		
<hr/>					
QY	133	gsaaataacgcgggtgcctaatacgytcgcggatgtgtgtatcgatgggggaacatgagaaaaa	192		
Db	226	agtccaacgcggggccaatatcgattcgtgcgtttgttcataatgvggaaaagaatgaccaaaa	285		
<hr/>					
QY	193	gatgatacccatacagtgaaagaaaccttactcttttcgcggaagatatcatcttggttgtc	252		
Db	286	gatgatcgaaaacatggtatgcttcgtgtattctcttgcgcagagacagatcaaaatgatgtgcc	345		

YY	253	gtctctgaattcatatcatctcccggttttgatcttcattgcttcgtctcaaacgctgcgtt	312		
Db	346	gttttagaattcatatcatctactactgtagcaataattctctccgattctgcgaagcgcgtg	405		
QY	313	gattatgtgattgaatgagaaagtctcttaattgtgaagaagaaagataccgttatattaat	372		
Db	406	gacctattgattgattgaaagaggtgttttgcgagggaagaagagaaatgtagatcaattat	465		
QY	373	attcgcgaatgaatggtgtgttcgttgaggagagagatgctgtggcgcgaaggtataaaaca	432		
Db	466	atcgcaatgtaaatggtacggtgtccttgaggcgaagcgcatggtgcacgagggtagtcagaa	525		
QY	433	gcaatcccgagatgtgcgtaaacgcgggtctaaacacataccattgatgtgatgtcgcggg	492		
Db	526	gcgatactgcagcttcgaatgacaggtctgtcatcacatattatgtttgacgtgcgcgt	585		
QY	493	tggggaaattccacaatgcatagtatctatgtttgtgaagaagaaagttttaatgctgcacct	552		
Db	586	tatggccagatcccatctcatctggtagttgattatggttcagaagatataatagctgcacca	645		
QY	553	caaggaataacaatggttttcgattcatatgataatgatgcaggtgtgaatgcatacgcaa	612		
Db	646	caggaaacaacaatgattctctgtcatatgtatgaaatgatgcagcgcgaatgcttaata	705		
QY	613	gttcgtactaattatgacgcgagttcttaatacagaacctgcattagttatggtgtaatt	672		
Db	706	gtaaagaaacaacattgctcgtcatcttaaaacccgaacttagctctgtatattgtgaatlc	765		
QY	673	ggaacccgtcatacaatggtgtgcgcgtgcgaatgaagcaacgatattagcatcttcgaaca	732		
Db	766	gggcattggtcattatgacggtgtgattgtatgtaggagacacatttaagctattcaagaa	825		
QY	733	agagagagttgggtgtgttcgcgtgtgtcaatggaaaggaacgcgcgaagaattggagattctta	792		
Db	826	agaatgtggagatggtttgcgttgagactgtgcattggaatagatgaaaggggtcgaatacttc	885		
QY	793	gaacttcgaatggtttggcttggaataaactacacgcttggaggaaatcacataatgtagt	852		
Db	886	gattatcgaatgaccttgcgtgttaatgcacgtcacatggtgtgggtgcatacgaatgtaaac	945		
QY	853	ggtccatattggtttaagagaacattcgcagattagacaccgatttt	897		
Db	946	gttcgaatggtgattcgtcaaacctctaaagaagagcagtggttt	990		
RESULT 9					
ID	245345				
XX	245345	standard; DNA; 960 BP.			
AC	245345:				
XX	27-MAR-2000	(first entry)			
DE	DNA encoding a <i>Bacillus</i> sp. mannannase enzyme.				
XX					
XX	Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;				
KW	endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;				
KW	mannan; galactomannan; glucomannan; galactoglucomannan; cellulose; fibre;				
KW	synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;				
KW	plant material degradation; recycled waste paper; paper making pulp;				
KW	guar; locust bean gum; mannan-containing food; coffee extract;				
KW	cleansing composition; machine washing; hard-surface cleaner;				
KW	dishwashing; oral; dental; contact lens; body-care composition;				
XX	fabric softener; oil well drilling; subterranean formation fracture; ss.				
OS	<i>Bacillus</i> sp.				
XX					
FH	Key { location/Qualifiers				
FT	CDS 1..960				
FT	/*tag- a				
FT	/product- "mannanase"				
FT	/note- "not termination codon given"				
FT	sig_peptide 1..84				


```

FT      /*tag- a
FT      /product= "mannanase"
FT      /note= "partial sequence"
XX
XX      WO9964619-A2.
XX
XX      16-DEC-1999.
XX
XX      10-JUN-1999; 99WO-DK00314.
XX
XX      10-JUN-1998; 98US-0111256.
XX      20-OCT-1998; 98DK-0001340.
XX      20-OCT-1998; 98DK-0001341.
XX      28-OCT-1998; 98US-0105970.
XX      28-OCT-1998; 98US-0106054.
XX      23-DEC-1998; 98DK-0001725.
XX      05-MAR-1999; 99DK-0000306.
XX      05-MAR-1999; 99DK-0000307.
XX      05-MAR-1999; 99DK-0000308.
XX      05-MAR-1999; 99DK-0000309.
XX      09-MAR-1999; 99US-0123543.
XX      10-MAR-1999; 99US-0123623.
XX      10-MAR-1999; 99US-0123641.
XX      11-MAR-1999; 99US-0123642.
XX
XX      (NOVO ) NOVO-NORDISK AS.
XX
XX      Kaupinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX      WPI: 2000-105891/09.
XX      P-PSDB; Y54133.
XX
XX      New mannanases for treatment of textiles, plant material and coffee
XX      extract, and in cleaning compositions
XX
XX      Disclosure; Page 229; 242pp; English.
XX
XX      The present sequence encodes a mannanase enzyme (also known as
XX      mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
XX      The mannanase hydrolyses galactomannans. Specifically, mannanases
XX      hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
XX      glucamannans, and galactoglucomannans. The mannanase protein, or
XX      preparations containing it, are used to improve properties of cellulosic
XX      or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX      sizes or printing pastes). They are also used to degrade or modify
XX      plant materials (particularly recycled waste paper, paper making pulps,
XX      or material containing guar or locust bean gums (thickeners), or to
XX      reduce viscosity of mannan-containing foods or feeds). The mannanases
XX      are also used to process coffee extracts (to inhibit gel formation); in
XX      cleaning compositions (for machine washing of fabrics, as hard-surface
XX      cleaners, for hand or machine dishwashing, also in oral, dental, contact
XX      lens or body-care compositions) where they remove mannan-containing
XX      soils and prevent binding of some soils to cellulose; and in fabric
XX      softeners. They can also be used in oil well drilling to fracture
XX      subterranean formations.
XX
XX      Sequence 564 BP; 159 A; 111 C; 158 G; 136 T; 0 other:
XX
XX      Query Match 12.2%; Score 167.8; DB 21; Length 564;
XX      Best Local Similarity 59.2%; Pred. No. 1.le-38;
XX      Matches 286; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```

```

DB      201 cccgcacatagcagaacacagggcggaacacagtgagaaatcgctctatccaatgagacaga 260
QY      183 atggaacaaagatgacatccatcacagtaagaaccttactctttaagcgaagataca 242
DB      261 atcggaacaaagatgacgctcctcgagcttcgcccgtgctgctgcacagaaacatag 320
QY      243 ttggtgtctgtctctgaagttacatgactacggtttgattccattgctgcctca 302
DB      321 gttgacacacgctgtgaaagtcacacgactgtcacagaagatgaaatcccgatgattaga 380
QY      303 tctgtgtgctgattatgattgaaatgagaagtgcttaattggaagaagataccgt 362
DB      381 taagcagctcgaattactgactgacgaatgctgctgttcttaaggggaagacacgagct 440
QY      363 cattataatattgcgaatgattggttgcgttggaaggagatgcttggtcagc 422
DB      441 aatatttaacttgcgaatgattgattgagggcgtgagagatgacacttgaggcaaac 500
QY      423 gtataacaaagcaatcccgagattggtgaacgcccgtcttaaacatcattgattaga 482
DB      501 atacgcacaaagcattcccgctgctgacagtgctgacctcgcctccacatacgtaataattga 560
QY      483 tgc 485
DB      561 tgc 563

RESULT 11
ID      245343
ID      245343 standard; DNA; 915 BP.
AC      245343;
XX
XX      27-MAR-2000 (first entry)
XX
XX      DNA encoding a partial Bacillus sp. mannanase enzyme.
XX
XX      Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX      endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
XX      mannan; galactomannan; galactoglucomannan; cellulosic fibre;
XX      synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
XX      plant material degradation; recycled waste paper; paper making pulp;
XX      guar; locust bean gum; mannan-containing food; coffee extract;
XX      cleaning composition; machine washing; hard-surface cleaner;
XX      dishwashing; oral; dental; contact lens; body care composition;
XX      fabric softener; oil well drilling; subterranean formation fracture; ss.
XX
XX      Bacillus sp.
XX
XX      Key 1.915
XX      CDS /*tag- a
XX      /product= "mannanase"
XX      /note= "partial sequence"
XX
XX      WO9964619-A2.
XX
XX      16-DEC-1999.
XX
XX      10-JUN-1999; 99WO-DK00314.
XX
XX      10-JUN-1998; 98US-0111256.
XX      20-OCT-1998; 98DK-0001340.
XX      20-OCT-1998; 98DK-0001341.
XX      28-OCT-1998; 98US-0105970.
XX      28-OCT-1998; 98US-0106054.
XX      23-DEC-1998; 98DK-0001725.
XX      05-MAR-1999; 99DK-0000306.
XX      05-MAR-1999; 99DK-0000307.
XX      05-MAR-1999; 99DK-0000308.
XX      05-MAR-1999; 99DK-0000309.
XX      09-MAR-1999; 99US-0123543.
XX      10-MAR-1999; 99US-0123623.

```


XX SQ Sequence 1461 BP; 475 A; 257 C; 340 G; 389 T; 0 other;

Query Match
Best Local Similarity 7.9%; Score 109.4; DB 10; Length 1461;
Matches 259; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

910 tctgatgaggaactctccgaacactcttattgatttgaaggtatgtatgaagatgtg 969
991 tatgtacatacaatccagcactgtctctctatgtatttgaataacatgcctacg 1050
970 actggaagatgact--gagcggaagcttggtgctgacagatgtgtcttaagga 1026
1051 tccgggtggaatttcagcagggagacatgacttcgaatgagtcgacgaatgtgt 1110
1027 agtcatctttaaagcggatattcaattgtcgtcaaatccacatacttactacgtt 1086
1111 actcaatcgttgaagcagatgctgtcggcaataatagctacat--ttgcataaa 1167
1087 attcaaatagctcttaccagcagaatagtagtaccagctttaaactgtcaaat 1146
1168 acagtgaaatcgaaactctgtatcctcaaaaacctgaatgaatgagccattcttg 1227
1147 tggggaagtgctgtaataagatgacgctgcttattgtgaaacagagacatgttat 1206
1228 tggggaatgtagaagtgtagcagcagagagtttcgtcaaaacagggagtgcttg 1287
1207 acatgtaactctggaagcttggtgcgattaaaggttcatctgcaacacgctactct 1266
1288 agatgaaatcaggtgaatttgcgtgttcgagcgaacacacgacgacatctcatt 1347
1267 gatttacaatctccaaactctctcaagtaaggaattggaattcagttccatca 1326
1348 gatttgcgaagtaagtaacttcgtcgtatgctgcgagagtaggtttaggtataagca 1407
1327 gcgagtgatagtgtagacaaacatcgattatataatgtatgtatgta 1377
1408 ccagcaaatagcaacggaagcagcgtattacttagatcatgtgacgta 1458

RESULT 13
245344 ID 245344 standard; DNA; 397 BP.

XX AC 245344;
XX DT 27-MAR-2000 (first entry)
XX DE DNA encoding a partial Bacillus mannanase enzyme.
XX KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture; ss.
XX OS Bacillus clausii.
XX XX
XX Key Location/Qualifiers
XX CDS 1..396
XX FT /*tag- a
XX FT //product- "mannanase"
XX FT //note- "partial sequence"
XX W09964619-A2.
XX PN 16-DEC-1999.
XX PD
XX XX

PF 10-JUN-1999; 99WC-DK00314.
XX 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105870.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 05-MAR-1999; 99US-0123543.
PR 09-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX (NOVO) NOVO-NORDISK AS.
XX KAupinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX WPI: 2000-105891/09.
XX P-PSDB: Y54131.
XX New mannanases for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions -
XX Disclosure; Page 226-227; 242pp; English.

XX The present sequence encodes a mannanase enzyme (also known as
XX mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
XX The mannanase hydrolyses galactomannans. Specifically, mannanases
XX hydrolyse 1,4-beta-D-mannosidic linkages in mannan, galactomannan,
XX glucomannans, and galactoglucomannans. The mannanase protein, or
XX preparations containing it, are used to improve properties of cellulosic
XX or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannanases
XX are also used to process coffee extracts (to inhibit gel formation); in
XX cleaning compositions (for machine washing of fabrics, as hard-surface
XX cleaners, for hand or machine dishwashing, also in oral, dental, contact
XX lens or body-care compositions) where they remove mannan-containing
XX soils and prevent binding of some soils to celluloses; and in fabric
XX softeners. They can also be used in oil well drilling to fracture
XX subterranean formations.

SQ Sequence 397 BP; 113 A; 69 C; 116 G; 99 T; 0 other;

Query Match
Best Local Similarity 7.6%; Score 104.4; DB 21; Length 397;
Matches 189; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

910 tctgatgaggaactctccgaacactcttattgatttgaaggtatgtatgaagatgtg 969
991 tatgtacatacaatccagcactgtctctctatgtatttgaataacatgcctacg 1050
970 actggaagatgact--gagcggaagcttggtgctgacagatgtgtcttaagga 1026
1051 tccgggtggaatttcagcagggagacatgacttcgaatgagtcgacgaatgtgt 1110
1027 agtcatctttaaagcggatattcaattgtcgtcaaatccacatacttactacgtt 1086
1111 actcaatcgttgaagcagatgctgtcggcaataatagctacat--ttgcataaa 1167
1087 attcaaatagctcttaccagcagaatagtagtaccagctttaaactgtcaaat 1146
1168 acagtgaaatcgaaactctgtatcctcaaaaacctgaatgaatgagccattcttg 1227
1147 tggggaagtgctgtaataagatgacgctgcttattgtgaaacagagacatgttat 1206
1228 tggggaatgtagaagtgtagcagcagagagtttcgtcaaaacagggagtgcttg 1287
1207 acatgtaactctggaagcttggtgcgattaaaggttcatctgcaacacgctactct 1266
1288 agatgaaatcaggtgaatttgcgtgttcgagcgaacacacgacgacatctcatt 1347
1267 gatttacaatctccaaactctctcaagtaaggaattggaattcagttccatca 1326
1348 gatttgcgaagtaagtaacttcgtcgtatgctgcgagagtaggtttaggtataagca 1407
1327 gcgagtgatagtgtagacaaacatcgattatataatgtatgtatgta 1377
1408 ccagcaaatagcaacggaagcagcgtattacttagatcatgtgacgta 1458

XX 4 gcaatccggaatttatgaagcgtgacactctatagatgcaatggaaccattt 63
XX 67 ggcgaacatgatttcaagtaacagcagcaggttgcgtgagggagcattcgat 126
XX 64 gtaatgagaggaattacaatggcgcagcatgtatataagaccaggcaactctcatt 123
XX 127 gtatgctgtgagtcgaatcagcagcattcatgttcaacagacactgtatcagcaata 186
XX 124 gaaggatgcaaatccggtgtaatacgtccgagatgtgtatcgtatgagggacaa 183
XX 187 ccagctattgcagcagcgtgcgttaatacgtgagatcgtttatcgaatgccaaca 246
XX 184 tggacaataatgacatccatcagtaagaaacattatctcttagcggaaataatcat 243
XX 247 tgggagagagatcagctagcgaagtggaagtgcttgcagttaccgaagaggaagc 306
XX 244 ttgtgtcgttcttgaagttatgtagtaccggttatgattcattcttgcgcaat 303

Matches	82: Conservative	0: Mismatches	67: Indels	0: Gaps
QY	806 attgagctcggaaataaactctacagcttggggaaatacatagtcgaatgctcatggtt	865		
Db	376 atggtcgtggaattatccactctcttagtgcgtatcatgaattatctccatagtat	435		
QY	866 taaggagaacttcggattaaagcacgcgttttccaggctggagatctgatatgggaactt	925		
Db	436 taatgcgtctcttattattatcatcgcgtctatcttaaggatggaatgacugtgctgcaact	495		
QY	926 ctccgacacactcttatgatttgaaagt	954		
Db	496 atcaagcacacagatttgattgaaagt	524		
RESULT 15				
XX	V74586/c			
XX	V74586 standard; DNA; 4826 BP.			
XX	V74586;			
XX	16-MAR-1999 (first entry)			
XX	Staphylococcus aureus contig SEQ ID #275.			
XX	Computer readable medium; vaccine; S.aureus infection; immunodetection;			
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;			
KW	skin infection; surgical wound infection; scalded skin syndrome;			
KW	toxic shock syndrome; ds.			
XX	Staphylococcus aureus.			
OS				
XX	Key	Location/Qualifiers		
XX	misc_feature	1081..1140		
XX		/tag= a		
XX		/note= "these bases represent a line of missing text in		
XX		the sequence listing in the specification. They		
XX		are included to maintain the nucleotide numbering		
XX		given in the specification for this DNA sequence"		
XX	misc_feature	2881..2940		
XX		/tag= b		
XX		/note= "these bases represent a line of missing text in		
XX		the sequence listing in the specification. They		
XX		are included to maintain the nucleotide numbering		
XX		given in the specification for this DNA sequence"		
XX	misc_feature	4681..4740		
XX		/tag= c		
XX		/note= "these bases represent a line of missing text in		
XX		the sequence listing in the specification. They		
XX		are included to maintain the nucleotide numbering		
XX		given in the specification for this DNA sequence"		
XX	EP786519-A2.			
XX	30-JUL-1997.			
XX				
XX	07-JAN-1997;	97EP-0100117.		
XX				
XX	05-JAN-1996;	96US-0009861.		
XX				
XX	(HDMN-) HDMN GENOME SCI INC.			
XX	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;			
XX	Rosen CA;			
XX	WPI: 1997-374922/35.			
XX				
XX	Polynucleotide(s) and proteins derived from Staphylococcus aureus			
XX	stored on computer readable medium and used in the production of			
XX	anti-S.aureus vaccines			
XX	Claim 1; Page 1089-1092; 3271pp; English.			

CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the *S. aureus* DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against *S. aureus* infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the *S. aureus* DNA sequences contained on the
 CC computer readable medium.

CC
 XX
 SQ Sequence 4826 BP; 1645 A; 707 C; 893 G; 1400 T; 181 other;

Query Match 3.0%; Score 41.8; DB 18; Length 4826;
 Best Local Similarity 55.0%; Pred. No. 0.045; Indels 0; Gaps 0;
 Matches 82; Conservative 0; Mismatches 67;
 QY 806 attggccttgaaataacacctacagcttggggaatacaatagtgatccatatggt 865
 DB 348 ATGGTCTTGAAATATCCACTCTTGTGGTATCATGAATTTATTCATATGTA 289
 QY 866 taagagaacttcgagattaagaccgctttacagggtgagagatcgtgaggaactt 925
 DB 288 TAATGCTCCTTATATTTATCGCTGTTATTAGTGAATGACTGTTGCAACT 229
 QY 926 ctccgaactcttattgatttgaagt 954
 DB 228 ATCAAGCAACAGGATTTGATTTAAAGT 200

Search completed: December 20, 2000, 09:59:18
 Job time: 30632 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:56:44 ; Search time 75.74 Seconds
(without alignments)
2735.645 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_1470
Perfect score: 1380
Sequence: 1 aatgcaatccgattta.....ttgataatgattgtagaa 1380

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*

- 1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/lna/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/1/lna/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/1/lna/6_COMB.seq:*
- 6: /cgn2_6/ptodata/1/lna/PCrUS_COMB.seq:*
- 7: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	48.6	3.5	7218	1	US-08-232-463-14
2	33.4	2.4	1365	4	US-08-870-827-4
3	33.4	2.4	2408	4	US-08-870-827-5
4	33	2.4	782	1	US-08-261-825-1
5	33	2.4	782	3	US-08-719-124-1
6	33	2.4	782	6	PCT-US95-07748-1
7	33	2.4	782	6	PCT-US95-07748-1
8	33	2.4	1800	5	US-09-039-773A-1
C 9	32.4	2.3	5467	1	US-07-745-206A-12
C 10	32.4	2.3	5467	3	US-08-311-363-12
11	32	2.3	4359	3	US-08-566-398-23
12	32	2.3	4365	6	PCT-US91-08525-21
13	32	2.3	4365	6	PCT-US91-08525-25
14	32	2.3	4365	6	PCT-US93-04384-1
15	32	2.3	4365	6	PCT-US93-04384-7
16	32	2.3	4365	6	PCT-US93-04384-15
17	32	2.3	6090	3	US-08-566-398-35
18	32	2.3	6144	7	US-08-566-398-32
C 19	31.8	2.3	7366	7	5169760-3
C 20	31.8	2.3	7852	5	US-08-836-022A-2
C 21	31.8	2.3	7897	5	US-08-836-022A-1
C 22	31.8	2.3	8509	2	US-08-462-014-1
C 23	31.8	2.3	10398	3	US-08-331-384-1
C 24	31.8	2.3	10398	3	US-08-708-188-1
C 25	31.8	2.3	10398	3	US-08-836-087-1
C 26	31.8	2.3	19182	4	US-08-850-880-11

C 27	31.8	2.3	19182	4	US-08-944-916-11	Sequence 11, Appl
C 28	31.6	2.3	40328	5	US-08-742-185-102	Sequence 102, App
C 29	31.6	2.3	43795	5	US-08-742-185-101	Sequence 101, App
30	31.4	2.3	8855	3	US-08-542-003-1	Sequence 1, Appl
31	31.4	2.3	8855	3	US-08-322-760A-1	Sequence 1, Appl
32	31	2.2	1453	4	US-08-252-995D-1	Sequence 1, Appl
33	31	2.2	1453	4	US-08-834-108-1	Sequence 1, Appl
34	31	2.2	1600	4	US-08-252-995D-5	Sequence 5, Appl
35	31	2.2	1600	4	US-08-834-108-5	Sequence 5, Appl
36	31	2.2	2882	1	US-08-424-567-1	Sequence 1, Appl
37	31	2.2	2882	3	US-08-711-928-1	Sequence 3, Appl
38	31	2.2	3447	1	US-08-252-995D-3	Sequence 3, Appl
39	31	2.2	3447	4	US-08-834-108-3	Sequence 4, Appl
C 40	30.6	2.2	18994	2	US-08-459-586-4	Sequence 4, Appl
C 41	30.6	2.2	18994	3	US-08-282-696-4	Sequence 4, Appl
C 42	30.2	2.2	5718	5	US-08-714-918-48	Sequence 48, Appl
C 43	30	2.2	2675	1	US-07-749-001-4	Sequence 4, Appl
C 44	30	2.2	2675	1	US-08-154-198-4	Sequence 4, Appl
C 45	30	2.2	2675	1	US-08-463-335-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: PC floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)836-4109
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-F1s
US-08-232-463-14

Query Match	3.5%;	Score 48.6;	DB 1;	Length 7218;
Best Local Similarity	3.8%;	Pred. No. 0.00011;		
Matches	15;	Conservative 216;	Mismatches 160;	Indels 0;
			Gaps	0

[illegible]

```

1 RESULT 2
2 US-08-870-827-4
3 ; Sequence 4, Application US/08870827
4 ; Patent No. 5962297
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Tsusaki et al.
7 ; TITLE OF INVENTION: POLYPEPTIDE HAVING
8 ; TITLE OF INVENTION: ACTIVITY
9 ; NUMBER OF SEQUENCES: 6
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: BROWDY AND NEWMARK
12 ; STREET: 419 Seventh Street, N.W., Suite 300
13 ; CITY: Washington
14 ; STATE: D.C.
15 ; COUNTRY: USA
16 ; ZIP: 20004
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Floppy disk
19 ; COMPUTER: IBM PC compatible
20 ; OPERATING SYSTEM: PC-DOS/MS-DOS
21 ; SOFTWARE: Patent In Release #1.0, Version #1.30
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/08/870.827
24 ; FILING DATE: 06-JUN-1997
25 ; CLASSIFICATION: A35
26 ; PRIORITY APPLICATION DATA:
27 ; APPLICATION NUMBER: JP 170,630/1996
28 ; FILING DATE: 10-JUN-1996
29 ; ATTORNEY/AGENT INFORMATION:
30 ; NAME: BROWDY, Roger L.
31 ; REGISTRATION NUMBER: 25, 618
32 ; REFERENCE/DOCKET NUMBER: TSUSAKI-2
33 ; TELECOMMUNICATION INFORMATION:
34 ; TELEPHONE: 202-628-5197
35 ; TELEFAX: 202-737-3528
36 ; INFORMATION FOR SEQ ID NO: 4:
37 ; SEQUENCE CHARACTERISTICS:
38 ; LENGTH: 1365 base pairs
39 ; TYPE: nucleic acid
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

STRANDEDNESS: double
TOPOLOGY: linear
US-08-870-827-4

	Query March	Similarity	2.44;	Score 33.4;	DB 4;	Length 1365;
	Best Local	Similarity	55.78;	Pred. No. 1.7;		
	Matches	64;	Conservative	0;	Mismatches	51;
					Indels	0;
						Gaps
0y	172 gatggggacatgacacaagaatgacacatcacagtaagaacctatctcttagcg	231				
Db	655 GATGAGGCAGTGGATTTCAGGTGATGATACCATCTTTAAAGAGCCCGCACTATGTGGAA	714				
0y	232 gaagataaactcttggttcgtcttcctgaagttcaatgactacccggttatgatt	286				
Db	715 GATTAAGGGCCATAATATCTGTCTTTTAAACGAGTAATCGAACAACAGATGCTT	769				

```

1  RESULT 3
2  US-08-870-827-5
3  Sequence 5, Application US/08870827
4  Patent No. 5862297.
5  GENERAL INFORMATION:
6  APPLICANT: Tsusaki et al.
7  TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
8  TITLE OF INVENTION: ACTIVITY
9  NUMBER OF SEQUENCES: 6
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: BROWDY AND NEIMARK
12 STREET: 419 Seventh Street, N.W., Suite 300
13 CITY: Washington
14 STATE: D.C.
15 COUNTRY: USA.
16 ZIP: 20004.
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: Patent In Release #1.0, Version #1.30
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/870,827
24 FILING DATE: 06-JUN-1997
25 CLASSIFICATION: 435
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: JP 170,630/1996
28 FILING DATE: 10-JUN-1996
29 ATTORNEY/AGENT INFORMATION:
30 NAME: BROWDY, Roger L.
31 REGISTRATION NUMBER: 25,618
32 REFERENCE/DOCKET NUMBER: TSUSAKI-2
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: 202-628-5197
35 TELEFAX: 202-737-3528
36 INFORMATION FOR SEQ ID NO: 5:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 2408 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: double
41 TOPOLOGY: linear
42 MOLECULE TYPE: genomic DNA
43 ORIGINAL SOURCE:
44 ORGANISM: Bacillus sp.
45 INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
46 FEATURE:
47 NAME/KEY: 5'UTR
48 LOCATION: 1..360
49 IDENTIFICATION METHOD: E
50 NAME/KEY: signal peptide
51 LOCATION: 361..456
52 IDENTIFICATION METHOD: S
53 NAME/KEY: mat peptide
54 LOCATION: 457..1821
55 IDENTIFICATION METHOD: S

```


NAME/KEY: 3'UTR
LOCATION: 1822..2408
IDENTIFICATION METHOD: E
US-08-870-827-5

Query Match 2.4%; Score 33.4; DB 4; Length 2408;
Best Local Similarity 55.7%; Pred. No. 2.2;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 172 gatggggagcaatggaacaaatgacatccatcacgaagaacctctctcttagcg 231
DB 1111 GATGAAGGCAAGTGGATTTCAGGTGATTAACACTTTAAGACACCTCATGTGGA 1170
QY 232 gaagataatcattggtctgtcttctgaagttcatgatcgctgctatgatt 286
DB 1171 GATAGGCCATTAATATCTTGTCTTGAAGCAATATCTGACACACAGATGCTT 1225

RESULT 4
US-08-261-825-1

Sequence 1, Application US/08261825
Patent No. 5558993

GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA

ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,825
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa

FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-261-825-1

Query Match 2.4%; Score 33; DB 1; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgtccaaatcttctccagtaagggaattggagttccatcagcgagtga 1334
DB 35 AAATTACCAAAATCTTAATCAACAAATGCTTAATTGCGAACACCAAAATAGCTATTGT 94
QY 1335 tagtagtgacaacacatcgattatattgataatgtagtctgtaga 1379
DB 95 TAGTCTGTCTTAATCTTAATAGTATTGCTAATAAGAAATGTATA 139

RESULT 5
US-08-719-124-1

Sequence 1, Application US/08719124
Patent No. 5854395

GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA

ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,124
FILING DATE: 24-SEP-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,825
FILING DATE: 17-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa

FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-719-124-1

Query Match 2.4%; Score 33; DB 3; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgtccaaatcttctccagtaagggaattggagttccatcagcgagtga 1334
DB 35 AAATTACCAAAATCTTAATCAACAAATGCTTAATTGCGAACACCAAAATAGCTATTGT 94
QY 1335 tagtagtgacaacacatcgattatattgataatgtagtctgtaga 1379

Db 95 TAGCTGTCTAATCTATATGTAATGCAATAGACTATA 139

RESULT 6
PCT-US95-07748-1
Sequence 1, Application PC/TUS9507748
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07748
FILING DATE: 16-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07419/013M01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
PCT-US95-07748-1

Query Match 2.4%; Score 33; DB 6; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgccaacatcttctcaagtaaggaaattgagttcgaatcagcgagtga 1334
DB 35 AAATTACCAAAATCTTAATCAACAATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1335 tagtagtgacaacacatcgattatattgataaagtgtgtgata 1379
DB 95 TAGCTGTCTAATCTATATGTAATGCAATAGACTATA 139

RESULT 7
PCT-US95-07748A-1
Sequence 1, Application PC/TUS9507748A
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07748A
FILING DATE: 16-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, Ph.D., LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07419/013M01 (PD3516)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
PCT-US95-07748A-1

Query Match 2.4%; Score 33; DB 6; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgccaacatcttctcaagtaaggaaattgagttcgaatcagcgagtga 1334
DB 35 AAATTACCAAAATCTTAATCAACAATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1335 tagtagtgacaacacatcgattatattgataaagtgtgtgata 1379
DB 95 TAGCTGTCTAATCTATATGTAATGCAATAGACTATA 139

RESULT 8
US-09-039-773A-1
Sequence 1, Application US/09039773A
Patent No. 6100388
GENERAL INFORMATION:
APPLICANT: Casas, Ivan
APPLICANT: Jonsson, Hans
APPLICANT: Wollstam, Bo
APPLICANT: Roos, Stefan
TITLE OF INVENTION: Lactobacilli Harboring Aggregation and Mucin
NUMBER OF SEQUENCES: 2
TITLE OF INVENTION: Binding Genes As Vaccine Delivery Vehicles
CORRESPONDENCE ADDRESS:
ADDRESSEE: Standley & Gilcrest
STREET: 495 Metro Place South, Suite 210
CITY: Dublin
STATE: Ohio
COUNTRY: US
ZIP: 43017
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS Version 6.22
SOFTWARE: Microsoft Word Version 6.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,773A
FILING DATE: 16-MAR-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA: NO. 6100388 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Donald G. Nickey
REGISTRATION NUMBER: 29,092
REFERENCE/DOCKET NUMBER: 1229-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 792-5555
TELEFAX: (614) 792-5536
TELEX: NO. 6100388 applicable
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Circular
MOLECULE TYPE: Genomic DNA
DESCRIPTION: Genomic DNA sequence and deduced amino
DESCRIPTION: acid sequence of bacterial aggregation
HYPOTHEICAL: No
ANTI-SENSE: Yes
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Lactobacillus reuteri sp
STRAIN: 1063
CELL TYPE: Unicellular organism
US-09-039-773A-1

Query Match 2.4%; Score 33; DB 5; Length 1800;
Best Local Similarity 53.5%; Pred. No. 2.5;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 507 acatcagattatgattggaagaagtttattatctaccctcaagaataacat 566
DB 1293 AGAAGCTTTTAAAGGCTCAAGTGCATCGCCCTTAATGATATCATGATTAATCGCGCA 1352
QY 567 gtctcagattcatatgtatgaatatgcaggtgtaatgcacgcgaattctactaatat 626
DB 1353 GGATTCACATCATCTGTTATGTAAGAACCCGTGAAGACCTATTAGAACTCATATGCAAC 1412
QY 627 tgaccgaagt 635
DB 1413 TGACCTAGT 1421

RESULT 9
US-07-745-206A-12/c
Sequence 12, Application US/07745206A
Patent No. 5429921
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: McCue, Ann
APPLICANT: Feldman, Daniel
TITLE OF INVENTION: Human Calcium Channel Compositions and
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitch, Even, Tabin & Flannery
STREET: 135 S. LaSalle
CITY: Chicago
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/745,206A
FILING DATE: 19910815
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Feder, Scott B
REFERENCE/DOCKET NUMBER: 51504
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-372-7842
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5467 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1c1n(144..3164, 3168..3245, 3249..3386, 3390
LOCATION: ..3392, 3396..3488, 3495..3539, 3543..3581, 3585
LOCATION: ..3587, 3591..3626, 3630..3689, 3693..3737, 3744
LOCATION: ..3746, 3750..4823, 4827..4841, 4845..5006, 5010
LOCATION: ..5096, 5100..5306, 5310..5366, 5370..5465)
US-07-745-206A-12

Query Match 2.3%; Score 32.4; DB 1; Length 5467;
Best Local Similarity 54.1%; Pred. No. 6.3;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 1185 tctgaacagagacatgttatctactctcgtgaagctttgtccgattaacgcttc 1244
DB 4884 TGCTAAGAGAGAGTACCGTACCTCTGCTGTGCTGATTAATAAATCGTGAACCTTC 4825
QY 1245 atctggaacacagctatctcagattatcaatgtaatgcacaaatcttccaagtaaggga 1304
DB 4824 ATCTGGGCAATGCAATCTCTGTACTAATAATATCATGTAATCTTCCCAACACAGTGA 4765
QY 1305 aa 1306
DB 4764 CA 4763

RESULT 10
US-08-311-363-12/c
Sequence 12, Application US/08311363
Patent No. 5876958
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: Feldman, Daniel
APPLICANT: McCue, Ann
APPLICANT: Brenner, Robert
TITLE OF INVENTION: Human Calcium Channel Compositions and
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,363

RESULT 11
 US-08-566-398-23
 Sequence 23, Application US/08566398
 Patent No. 5858373
 GENERAL INFORMATION:
 APPLICANT: Paoletti, Enzo
 TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIOUS
 TITLE OF INVENTION: PRIONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
 TITLE OF INVENTION: MAKING AND USING THEM
 NUMBER OF SEQUENCES: 63
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Curtiss, Morris & Safford, P.C.
 STREET: 530 Fifth Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/566,398
 FILING DATE: 01-DEC-1995
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:

	Query Match	2.3%;	Score 32;	DB 3;	Length 4359;	
	Best Local Similarity	65.3%;	Pred. No. 7.5;			
	Matches 47;	Conservative %	0;	Mismatches 25;	Indels 0;	
				Gaps	0;	
Qy	520	gattatggaagaagaattttaa	gctcaccctc	acacgaataat	acgaatttttcgattcat	579
Db	2587	gatttgcgaacaaatnccgttt	gaaatgcgtacaccct	agaatgtat	ttacaaatattgtttacacacaaatnt	2646
Qy	580	atgcatagaatat				591
Db	2647	ggtgtctgcacatg				2658

RESULT 12
 PCT-US91-08525-21
 Sequence 21, Application PC/TUS9108525
 GENERAL INFORMATION:
 APPLICANT: SmithKline Beecham, Corporation
 TITLE OF INVENTION: Recombinant feline Coronavirus S
 TITLE OF INVENTION: Proteins
 NUMBER OF SEQUENCES: 54
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SmithKline Beecham Corporation
 STREET: 709 Swedeland Road
 City: King of Prussia
 STATE: PA
 COUNTRY: USA
 ZIP: 19406-2799
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US91/08525
 FILING DATE: 19911114
 CLASSIFICATION: 435
 Prior Application DATA:
 APPLICATION NUMBER: US 07/698,927
 FILING DATE: 13-MAY-1991
 Prior Application DATA:
 APPLICATION NUMBER: US 07/613,066
 FILING DATE: 14-NOV-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: KING, William T.
 REGISTRATION NUMBER: 30,954
 REFERENCE/DOCKET NUMBER: SPC 14532B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 270-5015
 TELEFAX: (215) 270-5090
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4365 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: CDNA

ms-09-339-1_copy_91_1470.rni

Db 2653 GTGTCGATGT 2664

Search completed: December 20, 2000, 09:57:17
job time: 30824 sec

```

QY      580 atgtatgaaatat 591
          ||| ||| |||
DB      2653 GTGCTGCAATGT 2664

```

RESULT 15
PCT-US93-04384-7
: Sequence 7, Application PC/TUS9304384

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ADDRESS: Patents
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04384
FILING DATE: 19930507

1 CUBASIT APPLICATION DATA:
 2 PRIOR APPLICATION NUMBER: US 07/882,171
 3 FILING DATE: 08-MAY-1992
 4 PRIOR APPLICATION DATA:
 5 APPLICATION NUMBER: US 07/698,927
 6 FILING DATE: 13-MAY-1991
 7 PRIOR APPLICATION DATA:
 8 APPLICATION NUMBER: US 07/613,066
 9 FILING DATE: 14-NOV-1990
 10 ATTORNEY/AGENT INFORMATION:
 11 NAME: Schreck, Patricia A.
 12 REGISTRATION NUMBER: 33,777
 13 REFERENCE/DOCKET NUMBER: SBC H85009-1
 14 TELECOMMUNICATION INFORMATION:
 15 TELEPHONE: (215) 270-5015
 16 TELEFAX: (215) 270-5090
 17 INFORMATION FOR SEQ ID NO: 7:
 18 SEQUENCE CHARACTERISTICS:
 19 LENGTH: 4365 base pairs
 20 TYPE: NUCLEIC ACID
 21 STRANDEDNESS: double
 22 TOPOLOGY: unknown
 23 MOLECULE TYPE: cDNA
 24 FEATURE:
 25 NAME/KEY: CDS
 26 LOCATION: 1..4362
 27 PCT-US93-04384-7

Query Match	2.3%	Score 32;	DB 6;	Length 4365;
Best Local Similarity	65.3%	Pred. No. 7.5;		
Matches 47; Conservative	0;	Mismatches 25;	Indels 0;	Gaps 0;

QY 520 gattatggaagagaagtgttttaatgcctgaccctcaacgaatacactgttttggatccat 579
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2593 GATTGTCAAGATACGTTTGTATGGTAACCCTAGATGTAAACAATGTATAACCAATAT 2652
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 580 atgatgaatat 591
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 01:58:28 ; Search time 1967.85 Seconds
(without alignments)
4335.835 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_1470
Perfect score: 1380
Sequence: 1 aatgcaaatccgatttta.....ttgataatgtgattgtagaa 1380

Scoring table: IDENTITY_MDC
Gapop 10.0 , Gapext 1.0

Searched: 7189884 seqs, 3091403243 residues
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est4a:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

Db 954 KTAGAGGCKKKKKKDDRRDDAKATGKXTWRTTPTAKRRRRRRRBRAGRGDRKD 1013
Oy 972 tgaagtagcttgagcaggagctcttgctgacagagtgcttcttaagaagatca 1031
Db 1014 KKKRRARWRGKKKKKTDDDDGKRWGRAKAKADADADAADAAAWAKAKADK 1073
Oy 1032 ttctttaaagcgagat 1049
Db 1074 DKAKKADADDRKDKAK 1091

RESULT 2
CNS00396/c 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR08K10 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org/TheBDGP/Drosophila>
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_id="RPI-98"
/clone="BACR08K10"
/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 3.3%; Score 46.2; DB 121; Length 1101;
Best Local Similarity 14.5%; Pred. No. 0.05;
Matches 65; Conservative 215; Mismatches 164; Indels 3; Gaps 1;

Oy 683 atacaatggtgagctgacgaagcagatgatgagctatctgaacaagagagtg 742
Db 1026 WMMWATAKTTATWMTWMTARADWAGRGAGKRRDRADADAGRRDGGRRKKRDKD 967
Oy 743 ggtggtggtggtgacgaagagagagagagagagagagagagagagagagag 802
Db 966 RKDDDDDKGGKKKAAATATKMDMDKMDKMDGADRDADDGDDGGDDGGK 907
Oy 803 atgattggtggtgacatctacagcttggtggaatacagatgagtgccatag 862
Db 906 GKADDDDFDGRKDDDDKMDMDKAKGTWGDATWMAATWMMGMDADMDWTDAAAD 847

Oy 863 gtttaaggaacttgcga---gattaagcacgcgttttacaggtgagagatctgag 919
Db 846 DMMADDDDDAMWKKDDMMAMGARTADRRDWDGDRAGKGGAKRRDRKADKRAADDR 787
Oy 920 gaacttcctcgcaactcttatgatttgagagtagatgaagagagagagagag 979
Db 786 DDAATWTTWTTTTRDTRDMMKKTWTRMAADRWDADDDDRRAGTAGKMRTRKR 727
Oy 980 gcttgagcggaggtcttggtgacagagtggtcttcttaagaagagtcattaa 1039
Db 726 RWRRRDTRWDADADDTARDRRRRGGDAGAGKGTGRKRRRRATWTRTDAMWADA 667
Oy 1040 aagcgatattcaattgcgtcaaatcacacattacttaccatttaccataagct 1099
Db 666 AMWTTTDDTDWDRDRRRKRGARRRRTTAAADMTWAMDAKMDKTRADRWDRW 607
Oy 1100 cttaacagcagatagtagatcaag 1126
Db 606 AADTTDARKADRWAKARARARRDR 580

RESULT 3
CNS00LT2 1101 bp DNA GSS 14-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR48P19 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL078714 GI:5102004
KEYWORDS GSS.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org/TheBDGP/Drosophila>
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_id="RPI-98"
/clone="BACR48P19"
/note="end : TET3"

BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN

Query Match 3.1%; Score 43.4; DB 121; Length 1101;
Best Local Similarity 14.9%; Pred. No. 0.3;
Matches 48; Conservative 140; Mismatches 135; Indels 0; Gaps 0;

Oy 225 tttagcgaagaatacatcttggtgctgcttctgaagttcatgctaccggtatga 284

```

Db 699 TTTKKKKKAAAADKWDAAKATKTKKKDKKAAAMAAAKTKRKGKGGKGG 758
Qy 285 ttccattgcttcgctcaatgccttgcattatgattgaaatgagaagtgctta 344
Db 759 KKGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 818
Qy 345 tggaaaggaagatcgccttatatttgaaatgattgattgcttgcgggaag 404
Db 819 KKAARTRTKTKMDAAAAAAAKTKDKGKKKKTKTKTKKKKKKKKKGGKGA 878
Qy 405 ggaatgcttgctgcagcgtataaacaagcaatcccgatgctgaagcgcgt 464
Db 879 KKKGGTKKKKKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 938
Qy 465 ccatccttgatgctgctgcggggtgcggagcaattcccaatcattcat 524
Db 939 KGGKGGKTKTKTKKKKKKKKADAAAGKTKKRAAADAADWTADATKKKK 998
Qy 525 tggaaaggaagatttattgctg 547
Db 999 KTKKKDAAAATKTKKKKKKKTK 1021

```

RESULT 4

```

LOCUS CNS0182P 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
ACCESSION AL108811.1 GI:5629115
VERSION AL108811.1
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

```

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

```

source 1..1101
         /organism="Drosophila melanogaster"
         /plasmid="pBelobAC11"
         /db_xref="taxon:7227"
         /clone_1lb="DrosBAC"
         /clone="BACN37D10"
         /note="end : SP6"
BASE COUNT 274 a 268 c 128 g 73 t 358 others
ORIGIN

```

Query Match 3.1%; Score 43.4; DB 121; Length 1101;
Best Local Similarity 16.1%; Pred. No. 0.3; Mismatches 169; Indels 0; Gaps 0;

Qy 578 atatgatgaatagcagggtgtaatgacgcgaagttcgtactaataatgacgagtc 637
Db 692 ADARKAAAAAAGAAAAAAGAGGKRWKKBGDKGKATKTAAMAARXKMGTA 751
Qy 638 ttaataagactgcattgattgattgattgacacgcgtacatacaaatggtgacg 697

```

Db 752 WWTWATWADTWKATTDIDAKAAAGRRKRDARKTARGGRRATRRRMAAGRRAG 811
Qy 698 tgcataagcaacgattatgactatctgaacaagaagagtggtgctggtg 757
Db 812 AARAAARAAADDDRWDMWMAAAAAAAATTTWRBMDMDMDTDMDDTTAAWMD 871
Qy 758 catgaaagggagcggccagaatgagattttagaccttcgaatgattggtgaa 817
Db 872 DARAARRRRRRRRRRRARRARRAADDTDKRWADATDKPTTKTWTDDDDWDKAK 931
Qy 818 ataactacgctggggaataacatagatgctccatagtttaagaagaact 877
Db 932 KDRWMAARADGAKWKRDRARDMAATARDGDKMGKRGKRGKDKRWKDKTKG 991
Qy 878 cgaattagacacgcctttatcagtgagatgctgattgaggaagctccgcaact 937
Db 992 DDDWKTWTBMMWMTTKMDWMDGGRGRTWRKRWGAWRAADARDDTGKDTRT 1051
Qy 938 ttatgatttgaaggtatgatacgaagatgactggaagtagctgag 986
Db 1052 ADKRRRTDTYKRDGDWRKDRKDKRRKDKDKKATATDWDAD 1100

```

RESULT 5

```

LOCUS CNS01412 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
ACCESSION AL104180.1 GI:5615791
VERSION AL104180.1
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

```

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

```

source 1..1101
         /organism="Drosophila melanogaster"
         /plasmid="pBelobAC11"
         /db_xref="taxon:7227"
         /clone_1lb="DrosBAC"
         /clone="BACN1119"
         /note="end : SP6"
BASE COUNT 380 a 159 c 215 g 262 t 85 others
ORIGIN

```

Query Match 3.1%; Score 42.2; DB 121; Length 1101;
Best Local Similarity 40.8%; Pred. No. 0.63; Mismatches 150; Indels 1; Gaps 1;

Qy 985 agcggagtccttggtggtgacaggtgctctcaagaagatcattcttaagcg 1044
Db 722 AGCCCAAGCTTWWGGTTCAWMAACAAAACCCCAAAATGAGGTGGGAAATGATAACGC 791
Qy 1045 gataatcattgctcaaatcaacattactactatcatgttatcaaatagcttta 1104

```

Db 782 GAAAGCTGTAG-CCTTAAAGCTCGATTTKGCATGCCAGACACAAAATTCATGAC 840
Qy 1105 cagcagaatagtaggatacaagctactgttaacatgcaaatgtggagatgttggtaac 1164
Db 841 CTCACGACMACWAGAAAAGAACACCATWAGTATGATGATGAGAAAAGATGCAAAW 900
Qy 1165 ggaatgactgcgcgtcttattgtgaacacaggaatgttatatactgtactctggaagc 1224
Db 901 TMAATTTTGTGNTTARGTCGTAMAMATATWATDTGTTGTGCGACACACAAAGSAC 960
Qy 1225 ttgttgcgacttaacggttactctggaacacagctactctctagattatcaaatgtcaca 1284
Db 961 TGGGWTWAMACRKGACACAGCTAGRAGASGCTTACTGWCACATTAAGKCTCCMA 1020
Qy 1285 aatcct 1290
Db 1021 AMAGTT 1026

```

```

RESULT 6
LOCUS A2028169 471 bp DNA GSS 25-FEB-2000
DEFINITION RPCI-23-31608.TV RPCI-23 Mus musculus genomic clone RPCI-23-31608,
ACCESSION A2028169
VERSION A2028169.1 GI:7103553
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 471)
AUTHORS Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Aklnret,
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and and Fraser,C.M.
COMMENT Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-31608.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tcldb/Bac_ends/mouse/bac_end_intro.html
Plate: 316 row: 0 column: 8
Seq primer: T7
Class: BAC ends.

```

FEATURES

```

source
Location/Qualifiers
1..471
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-31608"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1:
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 151 a 115 c 88 g 117 t
ORIGIN

```

```

Query Match 3.0%; Score 41.8; DB 114; Length 471;
Best Local Similarity 56.0%; Pred. No. 0.63;
Matches 79; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
Qy 1210 tggactctggaagcttggcgcatgaaggttcatctcgtgaagaacgctatctgat 1269
Db 389 TGGTTGGTGGAGCTTTTGTGTTGCTTTAAACCTGACAGTATTCCTCAAAAT 330
Qy 1270 ttatcaatgtccaaatcttctcaagtaaggaattgagttcagttccaatcagcg 1329
Db 329 TCTGAAAATATTCAAAAGCTGAGAGTATGGAATTTGAGAGGAGTCCCTTTTCA 270
Qy 1330 agtatagtatgtgacaaaca 1350
Db 269 AGTCAGAGCCCTAAACAACA 249

```

```

RESULT 7
LOCUS C90049 350 bp mRNA EST 20-APR-1998
DEFINITION C90049 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
ACCESSION C90049
VERSION C90049.1 GI:3059669
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 350)
AUTHORS Yoshino,K., Morio,T. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1997)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402nuesakura.cc.tsukuba.ac.jp.
Location/Qualifiers
1..350
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSG140"
/clone_1lb="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 174 a 49 c 31 g 96 t
ORIGIN

```

```

FEATURES
source
Query Match 3.0%; Score 41.2; DB 36; Length 350;
Best Local Similarity 49.5%; Pred. No. 0.84;
Matches 106; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Qy 1047 tattcaattgctcaaatccaaatcttactacatgattatcaaatcgcgtttaca 1106
Db 307 TATTATTATTGTTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 248
Qy 1107 ggcgaataagtagatacaagctactgttaacatgcaaatgtggagatgttggtaag 1166
Db 247 GTAAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 188
Qy 1167 aatgactgcgcgtcttattgtgaacacaggaatgttatatactgtactctggaagct 1226
Db 187 ATGATATTATTCCTGATCTTGGAACTGGAACAATTAATTAATTCGAATTTGAAGCT 128
Qy 1227 ttgtcgaattaaaggttcatctggaacacagcta 1260
Db 127 GCTTCTATATTCATTTCAATGATGATGATGATGATGATGATGATGATGATGAT 94
RESULT 8

```

VERSION	KEYWORDS
SOURCE	ORGANISM
REFERENCE	AUTHORS
BB386799.1	GI:9111642
EST.	house mouse.
Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
(bases 1 to 303)	
Konno H., Aizawa K., Akahira S., Akiyama J., Arakawa T., Carinci P., Endo T., Fukuda S., Fukuishi Y., Hara A., Hayatsu N., Hirozane T., Horii F., Ishii Y., Ishikawa J., Ishikawa T., Itoh W., Izawa M., Kadota K., Kagawa I., Kai C., Kawai J., Kikuchi N., Kiyosawa H., Kojima Y., Kondo S., Koys S., Kurihara C., Kusarabe M., Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H., Okazaki Y., Ono T., Owa C., Saito H., Sekai C., Sato K., Shibata K., Shiba Y., Shigemoto Y., Shinagawa A., Shiraki T., Sogabe Y., Suganara Y., Suzuki H., Suzuki H., Tagawa A., Takahashi F., Tomiなが N., Toya T., Tsunoda Y., Watabiki A., Watanabe S., Yamamura T., Yamataka I., Yao R., Yasunishi A., Yokota T., Yoshida K., Yoshihi A., Yoshimura M., Muramatsu M. and Hayashizaki Y.	
Riken Mouse ESTs (Konno H., et al.)	
Unpublished (2000)	
Contact:	Yoshihide Hayashizaki

```

FEATURES
source

1. 303
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230058N01"
/clone_lib="RIKEN full-length enriched, 0 day neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site.1: SalI; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGAGATCCAGAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot - 20.0 and subtraction to Rot - 479.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTATTAATTATATCCCCCCCCCTT
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pluescript KS(+) after bulk excision from lambda280
FLC I."
```

LOCUS	A0621653	520 bp	DNA	GSS	16-JUN-1999
DEFINITION	HS_3089_B2_H12_MF	CIT	Approved	Human Genomic Sperm Library	D Homo

ACCESSION	sap1ens genomic clone Plate=3089 Col=24 Row=P, DNA sequence.		
VERSION	AO621653		
KEYWORDS	AO621653.1 GI:5084045		
SOURCE	GSS.		
ORGANISM	human.		
TITLE	Homo sapiens		
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
MEDLINE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 520)		
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,K., Furlong,J., Young,J., Zhao,S., Adams,N.D. and Hood,L.		
COMMENT	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999) 99380589 Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.hnsc.washington.edu Plate: 3089 row: P column: 24 Seq primer: T7 Class: BAC ends High quality sequence stop: 520.		
FEATURES	Location/Qualifiers 1..520 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_plate="3089 Col=24 Row=P" /clone_id="CIT Approved Human Genomic Sperm Library D" /sex="male" /note="Organ: sperm; Vector: pBelBAC11; BAC clones in E-Coli DH10B"		
BASE COUNT	146 a 100 c 129 g 143 t 2 others		
ORIGIN			
Query Match	2.8%; Score 39.2; DB 100; Length 520;		
Best Local Similarity	57.3%; Pred. No. 3.4;		
Matches	71; Conservative 0; Mismatches 53; Indels 0; Gaps 0;		
OY	223 tctttagcggaagaatcaatcattgtgtctgtcttcttgaaatcatcatgtacacggtat 282		
Db	211 tctttctctcattacacagacttcttatttttttttttttttcgttgaatgtctccaggttt 270		
OY	283 gattccattctgtctgcataatcgtctcttcttatttttgattgaatagagatgcttta 342		
Db	271 gttttaaattggatgtttcaggaatttcggtttcaggaattgcactaaagagatgacattt 330		
OY	343 attg 346		
Db	331 aatg 334		
RESULT 13			
CNS016E2/c			
LOCUS	CNS016E2 1204 bp DNA GSS 26-JUL-1999		
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC		
ACCESSION	AL106628		
VERSION	AL106628.1 GI:5622852		
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		

REFERENCE	(bases 1 to 1204)
AUTHORS	Genoscope.
TITLE	Direct Submation
JOURNAL	Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : segreif@genoscope.cns.fr)
COMMENT	<p>Web : www.genoscope.cns.fr</p> <p>Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.biol.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.</p>
FEATURES	<p>Location/Qualifiers</p> <p>1..1204</p> <p>/organism="Drosophila melanogaster"</p> <p>/plasmid="pBelobAC11"</p> <p>/db_xref="taxon:7227"</p> <p>/clone_lib="DrosBAC"</p> <p>/clone="BACN15A12"</p> <p>/note="end : 17"</p>
BASE COUNT	298 a 172 c 106 g 316 t 312 others
ORIGIN	
Query Match	2.8%; Score 39.2; DB 121; Length 1204;
Best Local Similarity	19.6%; Pred. No. 4.4;
Matches	65; Conservative 134; Mismatches 131; Indels 2; Gaps 1
Qy	1045 gatactcaatgtcgcgaattccacaacttacttacatggtatccaaatagctcttta 1104
Db	1181 KATTAATKMMWAKKAKKAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAADA 1122
Qy	1105 cagcagaatag--tagatacaagctcactgttaacaatgacatggaatgggaagtgtgta 1162
Db	1121 AADAAAAAKKAKKAKKAAAAAADAAMKKKAAAAAATAAAAAAATAAAAAAATAAAAAA 1062
Qy	1163 atggaatgactgcgcgtcttatgtgaaacagcagcatggtatatacatgactctggaa 1222
Db	1061 DAKKKKKKAAADADKKAADKDKKAAADAMDADAADAAKKKKKKKDKKDDATDADAAA 1002
Qy	1223 gctttgcgcgataacgtttcactcggagacaacgcctactctagattatcaaatgtcc 1282
Db	1001 KAAKKKKKKKMMKKKKKKKAKKAKKAKKDKKDKKDKKDKKDKKDKKAKKAKKAAKDA 942
Qy	1283 aaaaattcttcagaagaaggaaatgagatcagttccaatcagcagagtatagtagtg 1342
Db	941 AAAAAAATKKKKKDAKAAKAKKKKKKAAKKKKDDAMADADKAAKAKKAAKAKKKDK 882
Qy	1343 gacaaacatcgattatattgataatgattc 1374
Db	881 KTDAAADKKAADAKKKKKDKAAKAAADKDAKK 850
RESULT 14	
LOCUS	AZ138670 750 bp DNA GSS 06-JUN-2000
DEFINITION	SP_01770_A2.C04.SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-170 Col-8 Row-E, DNA sequence.
ACCESSION	AZ138670
VERSION	AZ138670.1 GI:8290573
KEYWORDS	GSS.
SOURCE	Strongylocentrotus purpuratus.
ORGANISM	Strongylocentrotus purpuratus
REFERENCE	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus.
AUTHORS	Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Blondi,T.R., Swartzel,S., Wallace,V.C., Poustis,A.V., Livingston,B.T., Wray,G.A., Etlenschnn,C.A., Lehnach,H., Britten,R.J., Davidson,E.H. and

Wed Dec 20 15:35:13 2000

us-09-339-159-1_copy_91_1470.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 03:12:31 ; Search time 4256.96 Seconds

(without alignments)
923.387 Million cell updates/sec

Title: us-09-339-159-1_copy_91_990

Perfect score: 900
Sequence: 1 aatgcaaatccgagatttta.....gattaagcaccgtttttaca 900

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 segs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sy:*
27: em_un:*
28: em_un:*
29: em_vl:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pr4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pr5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_vl1:*
73: gb_vl2:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pr6:*
92: gb_pr7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	422.4	46.9	1551	1 AB007123	AB007123 Bacillus
2	239.2	26.6	1530	2 D86329	D86329 Vibrio sp.
3	134	14.9	4567	1 AF163837	AF163837 Caldicl
4	131.8	14.6	2155	2 CDCMANABD	M36063 C.saccharol
5	131.8	14.6	4977	2 CDCMANA	L01257 Caldocellum
6	125	13.9	5284	2 CS016508	U16308 Caldocellum
7	91	10.1	25970	2 SC2H4	AL031514 Streptomy
8	85.8	9.5	1881	74 STM8MANABE	M92297 Streptomyce
9	81.6	9.1	837	74 TF06227	AJ006227 Thermomon
10	52.4	5.8	14922	2 SCF73	AL121746 Streptomy
11	45.8	5.1	111071	53 AC024591	AC024591 Homo sapi
12	45.2	5.0	35049	33 CELK03E6	U55375 Caenorhabdi

13	45	5.0	175300	40	AC012459	Homo sapi
C	15	4.9	14235	73	H2067264	
C	14	4.8	784657	38	HS010859	
C	16	4.8	251206	68	AL354834	Human DNA
C	17	4.7	1063	72	HA067261	
C	18	4.6	214412	43	AC018695	
C	19	4.6	17660	48	SC9723	
C	20	4.6	38779	48	SC931900	
C	21	4.6	107642	12	AF125313	
C	22	4.6	165536	48	SC931900	
C	23	4.6	3734	8	AT8NAHE1	
C	24	4.6	81493	8	AT81KBEN	
C	25	4.6	82646	7	AB028611	
C	26	4.5	1302	11	AF090946	
C	27	4.5	56804	49	HS77N19	
C	28	4.5	2103	57	TM049024	
C	29	4.5	44101	60	AC073960	
C	30	4.4	14222	1	AE004401	
C	31	4.4	171766	41	AC016048	
C	32	4.4	2294	57	YMC09206	
C	33	4.4	11379	48	MTAJ2479	
C	34	4.4	187615	50	AC020724	
C	35	4.4	237523	2	RPX04	
C	36	4.3	2318	33	AF064521	
C	37	4.3	145011	33	AC009885	
C	38	4.3	147700	70	AP002362	
C	39	4.3	157591	43	AC018775	
C	40	4.3	164991	11	AC011088	
C	41	4.3	167492	70	AP000795	
C	42	4.3	169168	70	AP002497	
C	43	4.3	175297	70	AP001642	
C	44	4.3	2285	33	AF170854	
C	45	4.3	2290	33	AF044000	

ALIGNMENTS

RESULT	LOCUS	AB007123	1551 bp	DNA	BCT	28-APR-1998
AB007123	Bacillus	circulans	gene for	mannanase,	complete	cde.
AB007123	AB007123.1	GI:3090433				
AB007123.1	guar gum-degrading	enzyme;	mannanase.			
AB007123.1	Bacillus	circulans	DNA.			
AB007123.1	Bacillus	circulans				
AB007123.1	Bacteria;	Firmicutes;	Bacillus/Clostridium	group;	Bacillaceae;	
AB007123.1	Bacillus	1 (bases 1 to 1551)				
AB007123.1	Yoshida,S.					
AB007123.1	Submitted (08-SEP-1997)	to the	DBJ/EMBL/GenBank	databases.	Seq11	
AB007123.1	Yoshida, Research	Institute of	Technology, Konolke	Construction,		
AB007123.1	Bioengineering; 4-3-55	Denpo, Konohama-Ku,	Osaka 554, Japan			
AB007123.1	(Tel:06-461-0262, Fax:06-468-3659)					
AB007123.1	2 (sites)					
AB007123.1	Yoshida,S., Sako,Y. and	Uchida,A.				
AB007123.1	Cloning, sequence	analysis, and	expression in	Escherichia coli	of a	
AB007123.1	gene coding for an	enzyme from	Bacillus circulans	K-1 that	degrades	
AB007123.1	guar gum					
AB007123.1	Biosci. Biotechnol. Biochem.	62 (3), 514-520	(1998)			
AB007123.1	9823274					

FEATURES

CDS

Location/Qualifiers

1..1551

/organism="Bacillus circulans"

/db_xref="taxon:1597"

1..1551

/function="guar gum-degrading enzyme"

/codon_start=1

/transl_table=11

/product="mannanase"

/protein_id="BAA25878.1"

QY	2	atgcaaatccggaatttattgtaagcgggtaccactcttaacgaatgcaatggaacccat	61	/db_xref="GI:3090434"
QY	98	atgctgcaagcggattttatgtatgacgggtaccgaattttggtatgacgacgaacccat	157	/translation="MGWFLVILRMWLAFAVFLMFSWTGQITRKRAHAAAGFYSGTK
QY	62	ttgtaatggaaggatttaacacggtgcaacggtgtaaaagacgaaggcaactactgaa	121	LIDATGPFVKNRVNHTATWTKDLSIPATGANTITVILANGKRTLDVNTV
QY	158	ttgtatggaaggatgacatgacgacacacgacacgacacgacacgacacgacacgac	217	NNIILICEONKLIATVEVDATGDSLSDIADNATVNTGIGSLAKEDRIVYIANE
QY	122	ttgaagggaattgcaaatccggtgtaacggtgcaacggtgtaaaagacgaaggcaac	181	NYGTWDGVAMANGYKQALPKLRNMGFTLTVDSAGMGQYDSKNTGTEVLMADPK
QY	218	taccagcatttctgtaaaacaggtgcaacggtgcaacggtgtaaaagacgaaggcaac	277	NTVPSIHMYEYAGNATGNTGIVTKNLALITGEGCGGTGNDVDENTATMSQE
QY	182	atggaagcaagatgacatgacatgacatgacatgacatgacatgacatgacatgacat	241	KGVGTLAMSKNGSSDLALIDMTDMMKNSLSTGNTVYNSNGIKATSVLSLFGV
QY	278	atggaagcaagatgacatgacatgacatgacatgacatgacatgacatgacatgacat	337	TPTSFSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTST
QY	242	atggttgctgcttcttgaatgacatgacatgacatgacatgacatgacatgacatgac	301	EMKATGAQTLADVSLQSNSTHSLYSINOMLSGKSLKATVKANNGNIGNGIYAKL
QY	338	aaactaatgcccgttttggaagatgacatgacatgacatgacatgacatgacatgac	397	YKIGSGWTWYDSGENTLIQSDGTILTLISLGIINLSSEKIVEFEFASNSGSAI
QY	302	atcgctgctggtgatttattggaatggaagatgcttcaattggaaggaagataccg	361	YDVSYSIQ"
QY	398	acaaacgcccgttatttgcgaattggttatttgcgaattggttatttgcgaattggt	457	
QY	362	tcattatgaatgacatgacatgacatgacatgacatgacatgacatgacatgacat	421	
QY	458	taattatgaatgacatgacatgacatgacatgacatgacatgacatgacatgacat	517	
QY	422	ggtataaagcaatcccgatgacatgacatgacatgacatgacatgacatgacatgac	481	
QY	518	gttatgaagcaatcccgatgacatgacatgacatgacatgacatgacatgacatgac	577	
QY	482	atgctgctggtggtggaatgacatgacatgacatgacatgacatgacatgacatgac	541	
QY	578	actcgcgtggaatggaagcaatgacatgacatgacatgacatgacatgacatgacat	637	
QY	542	atgctgaccccaagcaatgacatgacatgacatgacatgacatgacatgacatgac	601	
QY	638	atgcaagacccgttataaaacacatgacatgacatgacatgacatgacatgacatgac	697	
QY	602	atgcaagcgaatgacatgacatgacatgacatgacatgacatgacatgacatgacat	661	
QY	698	atgcaagcgaatgacatgacatgacatgacatgacatgacatgacatgacatgacat	757	
QY	662	ttgtggaatggaagcgtgacatgacatgacatgacatgacatgacatgacatgacat	721	
QY	758	tcggcgaatttgggaacacatgacatgacatgacatgacatgacatgacatgacat	817	
QY	722	atgcaagcgaatggaagcgtggaatgacatgacatgacatgacatgacatgacatgac	781	
QY	818	attcccaagcgaatggaagcgtggaatgacatgacatgacatgacatgacatgacat	877	
QY	782	ggaggaatggaatggaatggaatggaatggaatggaatggaatggaatggaatggaat	841	

Query Match 46.9%; Score 422.4; DB 1; Length 1551;
Best Local Similarity 67.0%; Pred. No. 6.2e-91;
Matches 600; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

BASE COUNT 461 a 333 c 363 g 394 t
ORIGIN

Db	878	TGGCTTATTCGATTGACAAATGATGTGGCGGTGAAGCACTCCCTCACACTCCTGGGAANA	937
Oy	842	caataatgtaagtgctccatgatgtttaagaagaacttcgcgaattaaagcagctttt	897
Db	938	CCGTAGTGAAATGGCAGTAACGGCATTTAAAGCAACTCTGTGTATCCGGCATTTTT	993

RESULT	2
LOCUS	D66329 1530 bp DNA BCF 19-MAR-1998
DEFINITION	Vibrio sp. gene for beta-1,4-mannanase, complete cds.
ACCESSION	D66329
VERSION	D66329.1 GI:2978250
KEYWORDS	beta-1,4-mannanase; manA.
SOURCE	Vibrio sp. (strain:MA-138) DNA.
ORGANISM	Vibrio sp.
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
AUTHORS	1 (bases 1 to 1530)
JOURNAL	Tamaru,Y.
TITLE	Direct Submission
REFERENCE	Submitted (01-JUL-1996) to the DDBJ/EMBL/GenBank databases. Yutaka Tamaru, Mie University, Faculty of Bioreources, Utilization of Marine Products; 1515 Kamihama, Tsu, Mie 514, Japan (E-mail:s9tamaye@maric.blo.mie-u.ac.jp, Tel:0592-31-9561, Fax:0592-31-9557)
AUTHORS	2 (sites)
REFERENCE	Tamaru,Y., Araki,T., Morishita,T., Kimura,T., Sakka,K. and Ohmlya,K.
JOURNAL	Cloning, DNA sequencing, and expression of the beta-1,4-mannanase gene from a marine bacterium, Vibrio sp. strain MA-138 J. Ferment. Bioeng. 83: 201-205 (1997)
FEATURES	Location/Qualifiers
SOURCE	1..1530
	/organism="Vibrio sp."
	/strain="MA-138"
	/db_xref="taxon:678"
-35_signal	32..37
-10_signal	53..58
RBS	136..141
gene	149..1336
	/gene="manA"
	149..1336
CDS	/gene="manA"
	/codon_start=1
	/transl_table=1
	/product="beta-1,4-mannanase"
	/protein_id="BAA25188.1"
	/db_xref="GI:2978251"
	/translation="MKPTKAFSLLPFWASCARGFVNSGYLYENGSAFKRIGITLHAHWYDRLKSVALSGIATANTVRVLNNGYTRMTKNVDSDVTNIILAAANNLIALAEHVDVTGGESSNAASLDSADAWIELEKNLLIQEDVVIINLSNEPGRNNNDVAWAVYDSAIORLRASAGINHRTIVDDAPNMGCQDFMLNNAOFNFNSDPKLNTFFSYAMYVSYSANSYNDRISSEFTNNGVLVIGEFPASTRGFDLVGDGSIMERSETLSLEYIGMSWAGPICOTTDDLDIVNNMNNNSYSTMGNVYLINGONGIKSTLTATVETPCGDCNCDDSGEXEPICSSANDPDGDGMENMSCIYQDSSDTAPNPAGYPICQDESSDPDGDMGWENNAMCC."

BASE COUNT	469 a	211 c	333 g	517 t
ORIGIN				

Query Match	26.6%	Score 239.2;	DB 2;	Length 1530;
Best Local Similarity	57.1%;	Fred. No. 4e-47;		
Matches 521:	Conservative 0;	Mismatches 373;	Indels 18;	Gaps 4;

Oy	3	tgcacaatcccgattatgttaacgcgttacccaactcatatagatgcgaatgtaaacaccatt	62
Db	202	TGCTCATGCGGGTTTATTATGTAATCAAATGCTCTTTATATGAAGCAATGTAAGTGCCTT	261
Oy	63	tgtaatgagaggattaaccaatgggcacgcacatggtataaagaacccagcacactatgcaat	122
Db	262	TAAATTTAGAGGATAAATCATGACATCACTCTGTATACAGACAAGTTAAAGCGTGCGCTT	321
Oy	123	tgaagagatttgtcaatatcccggtgctcaatacgcgtcccgagtgttatctgtatcgtatgggggaca	182

[illegible]

JOURNAL Appl. Environ. Microbiol. 66 (2), 664-670 (2000)
 MEDLINE 20120520
 REFERENCE 2 (bases 1 to 4567)
 AUTHORS Sunna, A., Gibbs, M.D. and Bergquist, P.L.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUN-1999) Biological Sciences, Macquarie University,
 North Ryde, New South Wales 2109, Australia
 FEATURES
 source Location/Qualifiers
 1. 4567
 /organism="Caldibacillus cellulovorans"
 /db_xref="taxon:74586"
 CDS
 1. 513
 /note="orf1"
 /codon_start=1
 /product="unknown"
 /protein_id="AAF22273.1"
 /db_xref="GI:6651326"
 /translation="TPTPTPPSAAPTPPSAGSLVVOYRAADTNAGDNLKPHFR
 VNRGTSVPLSELTRYWTVVDGKPOVFNCDMAOVGCSNVGSEVFKLSTGRTADY
 IEITTSAGSLAAGASSGDIOVRINKNDWNTYNEANDYSYDPTKSPADNNRYTLVR
 NGLIMGVEP"
 misc_feature
 1. 60
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 misc_feature
 61. 510
 /note="Region: cellulose-binding domain type IIb"
 /evidence="not_experimental"
 gene
 620. 3412
 /gene="mana"
 620. 718
 /gene="mana"
 sig_peptide
 620. 3412
 /evidence="not_experimental"
 CDS
 620. 3412
 /gene="mana"
 /codon_start=1
 /transl_table=11
 /product="multidomain beta-1,4-mannanase precursor"
 /protein_id="AAF22274.1"
 /db_xref="GI:6651327"
 /translation="MNRRLIARISGMLAMVLIAMLAAYPKPEPAEAGHGVFATRT
 YACYVDKGVHNGDGLMKNIPACDALISGNYOFPMFGLISNAGRHEILIPDG
 LCGPRASFGKQARTDWTTRLOPQATIVNNAKHPETWLYTRDMDPTQPL
 KMSDEPTFESQVNTPIINSSGPDAAEYVQVLPKQGRHIIIMQKSDPAFVN
 CSDYFSGSPPLAIEYRGEDTSATDMQKPOLRIYVTSQAVPLTELRKYWYTN
 TPTPTPPSAGSLAIEYRGEDTSATDMQKPOLRIYVTSQAVPLTELRKYWYTN
 STQAOYFCDNAOIGCSNIRAOYFSLQPVGADSYIELSTFGGSPAGGSGTQIR
 IETNMWYNETDMQSYNGAQTWGPRLIRNGVYVWETEPGGSSPTPIVTEP
 PTPPTSTPTPTPTPTPTPSGGPNLSVNTQGLVGHIPHAMYRDRLSLSLQIRSM
 GANARIVLSNGCRWTKIPASEVADIIISQARTLYRAVYLVHDTTGEDAAASMT
 TAVNRIELKNVLAQENFVINGNEPYGNNTQNTWTDRAVQALRNAGINNTIM
 VDAPWQDMSTFMDNAPLIFNADPQNLVSLHMGYDTALEVOSYISFVNRGL
 PLVIEGFHMSDSDPNQALVOYAKOVNILEFGMSNGSGGVEYEDMTWEPANSP
 TAWGTWRTNAGISTPT
 KPHRIYRGTSVPLSELTRYWTVVDGKPOVFNCDMAOVGCSNVGSEVFKLSTGRT
 ADYIEITTSAGSLAAGASSGDIOVRINKNDWNTYNEANDYSYDPTKSPADNN
 RYTLIRNGQIMVGEV"
 mat_peptide
 719. 3409
 /gene="mana"
 /product="multidomain beta-1,4-mannanase"
 misc_feature
 1349. 1480
 /gene="mana"
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 misc_feature
 1481. 1927
 /gene="mana"
 /note="Region: cellulose-binding domain type IIb"
 /evidence="not_experimental"
 misc_feature
 1928. 2032
 /gene="mana"
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 misc_feature
 2033. 2884

Query Match 14.9% Score 134: DB 1: Length 4567;
 Best Local Similarity 51.9% Pred. No. 5.5e-22;
 Matches 408; Conservative 0; Mismatches 360; Indels 18; Gaps 4;
 /gene="mana"
 /note="Region: beta-1,4-mannanase catalytic domain"
 /evidence="experimental"
 misc_feature
 2885. 2959
 /gene="mana"
 /note="Region: proline-threonine linker"
 /evidence="not_experimental"
 misc_feature
 2960. 3409
 /gene="mana"
 /note="Region: cellulose-binding domain type IIb"
 /evidence="not_experimental"
 CDS
 3894. 4400
 /note="orf3"
 /codon_start=1
 /product="unknown"
 /protein_id="AAF22275.1"
 /db_xref="GI:6651328"
 /translation="MGASGVFSPRAONGYGLKARCKRRCVGNLIGRLVVDDEI
 RFPGARKRGRSRRNRNGEONPESGVTVVQDDKRIRREINQGLFVNDDEI
 GPKRLITIAQRQARAKKQASDQPLKNVILHFIDQHPYGVGRMDPFKFLTNVS
 RSLITRY"
 BASE COUNT 1040 a 1276 c 1439 g 812 t
 ORIGIN
 QY 68 tgaaggattaccatgagcagcatggtataagaccagcaactactgcaattgaag 127
 DB 2070 TGGTGGATATATACCCCGCATCGTGTCCGACCGGCTAGTTCGCTTGCAAG 2129
 QY 128 ggaatgcaataccggtgtaatacgggtccggaatggtatctgtaaggggaacatgga 187
 DB 2130 GCATTCGTTCTCGGGGCGCCAAACGGGTGGGATCGTGTGACAAACGGTGGCGGTGGA 2189
 QY 188 caaaggatgatacatcagaagaaacctatctcttcaagggaagataa---tcaat 244
 DB 2190 CGAAGATTCGGCCAGCAAGATGACGACATTATTTCTCAGCGCGGACGCTCGGTATC 2249
 QY 245 tgggtctgtcttgaagttcatgatacctcgggttattgattccattgc-----tt 295
 DB 2250 GGGGGGTGCTCTGTAAGTTACACACACGACCGGCTACGGGGAAGACCGCGGCGTGT 2309
 QY 296 cgcacaatcgctgctgattatgattgaatgaagaagtgcttcaatggaagaagaag 355
 DB 2310 CGATGACGACGCGCGTCAATTATGATTGAACGAAACGTCGCGGCTCAGGAGA 2269
 QY 356 ataccgcatattataatattgcaatgaatgattggttcgtgggaagggaatgctggg 415
 DB 2370 ATTTCGTCATGTCATATATCGGCATGACCGGTACGGGAACAATTAATCAGAACTGG 2429
 QY 416 ctgacgggtataaacaagcaatcccgcatgctgcaagccgggtcctaaccataccttga 475
 DB 2430 TGACGAGACACGGGAGACGGGTGACGGCGTCCGGAACCGGGGATCAACAAATACATCA 2489
 QY 476 tggtaagatgctgcgggttggggac---aattccacaatgattcatgattatgaagaag 532
 DB 2490 TGGTGGACGCTCCGAATCGAGTGTGACGACTGCTTCACAGATCGGGGACACGCGCGGA 2549
 QY 533 aagtttaatgctgacccccaacgaatacaatgatttggatttcattatgaatatg 592
 DB 2550 CGATTTCAACGCGGACCCCAAGCACTGCTTTTAAATCCACATGTACGGG---G 2606
 QY 593 caagtgtaatgcatgcaagttcgtactaataatgacagagttcttaacgaagacctcg 652
 DB 2607 TGRACGACACGGCGGAGAGTGAAGTATATGAGTGTGCTGTAACACCGGGGTTCG 2666
 QY 653 catgaagcatggtgaatttggacaacgctcatcaaatggtgacgctcgatgaagaacga 712
 DB 2667 CGGTGTGATCGGGGATTTTGTATATGACACGACGCGGGATTCGAAACGACGCGCA 2726

OY	713	ttatgagccattctgcaacaaagaggaattgggttggttcgtcggtcatggaaggaaggaac	772
Db	2727	tcgtacatatatctcgaaacgatacaacattggcttcttcgcgtcgtcgagaccgtaaac	2786
OY	773	gcccaagaatggagagattatgaaccttcgaatgatgagtcggctggaataaccttaagctt	832
Db	2787	gagcggcggtgaatattttggacatggtgacgaacttcaatcgcaagcccgagcggt	2846
OY	833	ggggaa	838
Db	2847	ggggga	2852
RESULT	4		
LOCUS	CDCMANABD	2155 bp	DNA BCT 26-APR-1993
DEFINITION	C.saccharolyticum beta-D-mannanase (mana) gene, complete cds.		
ACCESSION	M36063	M37147	
VERSION	M36063.1	GI:144292	
KEYWORDS	1,4-beta-D-mannan mannanohydrolase; beta-D-mannanase.		
SOURCE	C.saccharolyticum DNA, clone lambda-NZP2.		
ORGANISM	Caldicellulosiruptor saccharolyticus		
REFERENCE	Thermonaerobacter group; Caldicellulosiruptor.		
AUTHORS	Lueeth,E., Jasmat,N.B., Grayling,R.A., Love,D.R. and Bergquist,P.L.		
TITLE	1 (bases 1 to 2155)		
JOURNAL	Cloning, sequence analysis, and expression in Escherichia coli of a		
MEDLINE	gene coding for a beta-mannanase from the extremely thermophilic		
COMMENT	bacterium Caldocellum saccharolyticum		
	Appl. Environ. Microbiol. 57, 694-700 (1991)		
	91247819		
	Draft entry and computer-readable sequence kindly submitted by		
	E. Lueethi, 03-JUN-1990.		
	Author address: E.Lueethi		
	University of Auckland		
	Department of Cellular and Molecular Biology		
	Private Bag		
	Auckland		
	NEW ZEALAND		
	email:SYSTEM@CELVAX.AUKUNI.AC.NZ.		
FEATURES			
source	Location/Qualifiers		
	1..2155		
	/organism="Caldicellulosiruptor saccharolyticus"		
	/db_xref="taxon:44001"		
	<1..685		
	/codon_start=2		
	/label="ORI"		
	/protein_id="AA072860.1"		
	/db_xref="GI:144293"		
	/translation="GSMTWFGFQAMSMQRYVAEYYTGGDKAGALLKRWYSIKSVYK		
	LNSGTPAIPSTLDMGSDPDMWNGTGYGNPLAHYKVVYDGDGITSALNALLYSRA		
	GTRKGVYDEPARKLAKEILLDRMKLRDEKGLSAPKRDYKRFPEQVYIIPMGWG		
	KMRGPDYIKSGVKRITDIRSKYKODPMPKLEAAVYKSGVPEFRIRHWAQDIAIVNA		
	TYELFSGNQ"		
	818..830		
	/note="mana ribosome binding site"		
	841..963		
	/product="beta-D-mannanase"		
	841..1881		
	/EC_number="3.2.1.78"		
	/codon_start=1		
	/translation="beta-D-mannanase"		
	/product="beta-D-mannanase"		
	/protein_id="AA072861.1"		
	/db_xref="GI:144294"		
	/translation="MRLEKTKIRKRWLSYLCTVVFLLNIFLTAATIIIPKGAATSDNG		
	VVKIDTSTLIGTINAHCMYRDRLDTALGIRSGMGSVRYVLSGIRMTKIPASEVANA		
	ISISRSIGFKAIIILEVHDTGYGEDAGSLAAGEVWEIKRSVLTGQNEDFVILINGG		
	NEPKNVSNOMMYNDTKNAIKALRDAGSKFRIINDAPNMGQDMSNTKADNAGSIMEAD		
	PLRNLYVSIHMYGYNTASKEAYEIKSVYDGLPLVIGEGEHOHTDDPDEAIVARA		
	KQYKIGLFMSMCGNSSVYGLDMAVNNMKNPPIPMGQWYITNAIGISSIPTPISTYTT		

Query Match	14.6%	Score 131.8:	DB 2:	Length 2155:
Best Local Similarity	51.7%:	Fred. No. 1.9e-21:		
Matches 407:	Conservative 0:	Mismatches 362:	Indels 18:	Gaps 4
68	tgaagaggaattacacatgagcagcagatgataaagaccagccaactactgcaattgaag	127		
Dd	998 TAAATGGAACCAATACGCACATTCGTGTACAGAGATAGACTTGATACGGCATTCGGT	1057		
Qy	128 ggaattcaaatccggtcctaatacagtcgcgattgtgttattctgatctgaggggacaattga	187		
Dd	1058 GAATTAGGTCATGGGGGTATGAACTCTGTGAGGGGTAGTCTGATGATAGTATGCCATGGA	1117		
Qy	168 caaagaatgacatccatacagtaagaagaacctatctctttagcggagaattatcatttg	247		
Dd	1118 CGAATATCCACCAAGTGAAGGCGCAAAATTTATTCATTGTCACGAACCTCTGGTTTA	1177		
Qy	248 ttgcgctt---ctgaagtcatagtatgcaccggttaigtatccattg-----ctt	295		
Dd	1178 AAGCTATTTATTTAAAGTGCACGCACACAGATATGAGAAAGATGGCGCAGCATGTT	1237		
Qy	236 cgctcaatcgtcgttgttattatgtatgaaatgagaagtgccttaatttggaaaggaag	355		
Dd	1238 CATGGCACAAGCAATGGAAATTTGGAAGGAATTAAGAGCGTATTTAGACGGTAACGAAG	1297		
Qy	336 ataccgtcaattatataattcgcgaatgaatgtttggttcgtgggaaggagatgcttgg	415		
Dd	1238 ATTTGTATTTATTAACAATTGGTATATGAGCCGTATATGGAACAATATCTATCAAACTGGG	1357		
Qy	416 ctgacgggtataaacaacaccccgatgctgaagccgggtctaaccatacttga	475		
Dd	1358 TTAATGACACGAAAGACCTATTAAGACCTTAGAGATGCAAGATTCAGCACAGATTA	1417		
Qy	476 tggtagatgctgcggggtgggagc---aattccacaatcgatcattcattatggaaag	532		
Dd	1418 TGGTGATGCGCCGCAACTGGGGTCAAGGATTTGGTAAATACATGAGAGATATGCCACA	1477		
Qy	533 aagttttaahgctgacccctcaagaaatacaaatgttttcgattcatatgtatgaatg	592		
Dd	1478 GCATATATGGAACCAATCCGCTGCCCAATTGGTATTTTCGATTCATATATGATG---GCG	1534		
Qy	593 caggtgtaatgacatcgcaagttgctactaatatataagaccgagttcttaatacagaacctg	652		
Dd	1535 TATACATATACCAAGCAGAGGTCGAAGAGTACATCAATCATTTGTTAAGGGTTAC	1594		
Qy	653 catlagcatcgttgtaatttgagacacgltcatatacaaatgttgtagctcgataagacaaga	712		
Dd	1595 CATGTGTTATTTGGAATAATTTGGACATCAACAGATAGGTGACCTGATGAAGAAGCTA	1654		
Qy	713 ttatagactattctgacaacaagaagagttgggtgggttgcgtgctcatgtaagaaaggaacg	772		
Dd	1655 TTGTACGATATCAAAACAGTACAAAGATAGAGATTATTTAGTTGGTCGTGTGTGCAAAAT	1714		
Qy	773 gcccaagaatggagatatttagacccttcgaaatgattgggtcgtgaaataacccttaacgctt	832		
Dd	1715 CGAGCTATGTTGGATTGTCACATGTTAAACATCTGGAGCCCAATATTCACAACTCCAT	1774		
Qy	833 gggggaag 839			
Dd	1775 GGGGACA 1781			

mat_peptide

964.1878

/product="beta-D-mannanase"

BASE COUNT

721 a 334 c 590 g 510 t

ORIGIN

CDCCMANA

4977 bp

DNA

26-APR-1993

LOCUS

Caldocellum saccharolyticum beta-mannanase/endoglucanase (mana)

DEFINITION

gene, complete cds.

ACCESSION L01257
 VERSION 1.01257.1 GI:144290
 KEYWORDS beta-mannanase; endo-1,4 beta-mannanase; endoglyucanase; manA gene
 SOURCE Caldocellum saccharolyticum (library: Lambda NP2) DNA.
 ORGANISM Caldocellulositriptor saccharolyticus
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Thermomicrobacter group; Caldocellulositriptor.
 REFERENCE 1 (bases 1 to 4977)
 AUTHORS Gbbbs,M.D., Saul,D.J., Luthi,E. and Bergquist,P.L.
 TITLE The beta-mannanase from Caldocellum saccharolyticum is part of a
 multidiomalin enzyme
 JOURNAL Appl. Environ. Microbiol. 58, 3864-3867 (1992)
 MEDLINE 93119139
 FEATURES
 1 Location/Qualifiers
 14077

BASE COUNT	1698 a	815 c	1265 g	1199 t
ORIGIN				

Query Match	14.6%;	Score 131.8;	DB 2;	Length 4977;
Best Local Similarity	51.7%;	Pred. No. 1.8e-21;		
Matches 407; Conservative	0;	Mismatches 362;	Indels 18;	Gaps 4

Oy	68	tgaagagggattaaccatcgggagcagcagfghataaagcccaagcaactacgcaattgaag	127
Db	998	TAATATGGAAACCAATACCCACCAATTGCTGTGTACAGAGTATAGACTTGATACGGCACTTCGCG	105
Oy	128	ggattgcaaatcacggtgctataacggtccggattggttatctgatgaggggacaatgga	187
Db	1058	GATTATAGTCATGGGGTATGAACTCTGTGAGGGTAGTGCCTGAGTAATGGTTACCGATGGA	111
Oy	188	caaaagatgacatccatcacgtaagaagaaaccttactctctttaggggaaagabaatcattg	247
Db	1118	CGAAGATACCGAACAAGTAGTGGCAATATTATATCATTTGTCCAGAAGCTTGATTTCA	117
Oy	248	ttgtcgtt-----ctgaagttcatgatgactaacggtttatgattcatgtg-----cct	295

Db	1178	AACTTATTTATTAAGACGACAGCAGACAGATATGGAAGATGGGGCAGCATGTT	1237
QY	296	cgcctcaacgcgtcgtctgttgatttggattgaaatgaaatgctttaaattgaaagaaag	355
Db	1238	CATTGCACACAGCAGGAGATATTGGAAAGAGATAAAGGGTATTAAACGGTAAACGAG	1297
QY	356	atacgcctcatattaatatattcgaaatgaaatgttggttcgtcgaggaaaggaatcctgg	415
Db	1298	ATTTTGTATTTATTAACCATTTGGTATGATAGCCCTATGGGACAAATACATCAAAACTGGG	1357
QY	416	ctgacggtataaacaagcaatccgcggaattgcgtaaacgcggtcttaaccaatccttga	475
Db	1358	TTTATGACACGAGAAACGCTATTAAGCACTTATGAGATGCAGAGATTCAAGACACGATTA	1417
QY	476	tgttaagtctcgtagggatgagac--aatctcaaatcgatcatgattatgaaagag	532
Db	1418	TGGTGGATGCGCCGAAGTGGGGTCAAGATTGGTCTAATCTATGAGAGATATATGCCAGA	1477
QY	533	aagttttaatgctgacccctcaagaataacaaatglttgcattatagtatgaatctg	592
Db	1478	GCAATATGAGAGAGAGATCCGCTCGCCGCAATTTGGTATTTTCGATTCATATGATG--CGG	1534
QY	593	caggttgtaatgcatcgcgaagttcgcacaaatgatgaccaggttcttaaccaagaccctg	652
Db	1535	TATTCATATACAGCAACGACAGTGGAGAGTACTCAATCATTTGTTGTAAGGGGTAC	1594
QY	653	cattagtcatttgttgaattttgaaacccgcgtacatacaaatgtgtgtcgtatgaaagaaaga	712
Db	1595	CATTGGTATTATGGAGATTTGGCAATTCACACACACAAATGTAACCCGTATGAAAGAGTA	1654
QY	713	ttatgagctatcttcgaaacaagaagagttcggtgtgttcggtgcgtcatgaaaggaacg	772
Db	1655	TTTGTCAGCTATGCAAAACACTACAAATATGATATTATTGTTGTTGTCGTGGTGGAAATT	1714
QY	773	gcccaagaatggaagtatttagacccttcgaatgattggctgtaataataacttaacgctt	832
Db	1715	CGAGCTATGTTGGGTATTGGACATGTTAAACAACATGGGACCCCAATAATATCCAACTCCAT	1774
QY	833	gggggaa 839	
Db	1775	GGGGACA 1781	

```

RESULT      6
LOCUS       CSU16308
DEFINITION  Caldocellum saccharolyticum endoglucanase/mannanase (celc)
ACCESSION   U16308
VERSION     U16308.1
KEYWORDS    GI:577827
SOURCE      Caldocellum saccharolyticum,
            Caldicellulosiruptor saccharolyticus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Thermomanaerobacter group; Caldicellulosiruptor.
REFERENCE   1 (bases 279 to 4374)
AUTHORS     Morris,D.D., Reeves,R.A., Gibbs,M.D., Saul,D.J. and Bergquist,P.
TITLE       Correction of the beta-mannanase domain of the celc pseudogene from
            Caldocellulosiruptor saccharolyticus and activity of the gene
            product on Kraft pulp
JOURNAL     Appl. Environ. Microbiol. 61 (6), 2262-2269 (1995)
MEDLINE     95314280
REFERENCE   2 (bases 1 to 5284)
AUTHORS     Morris,D.D.
TITLE       Direct Submission
JOURNAL     Submitted (25-OCT-1994) Daniel D. Morris, School of Biological
            Sciences, University of Auckland, Private Bag, Auckland 92019, New
            Zealand.
FEATURES
source      Location/Qualifiers
            1..5284
            /organism="Caldicellulosiruptor saccharolyticus"
            /db_xref="taxon:44001"

```


JOURNAL

COMMENT

Submitted (01-SEP-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrelli@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>)
COS are numbered using the following system eg SC2H4.01c, SC (S. coelicolor), 787 (cosmid name), .01 (first COS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each COS.
Usually the highest scoring match found by fasta -o is given for COS which show significant similarity to other COS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid 2H4 lies between 265 and AH10 on the AseI-B genomic restriction fragment.
Location/Qualifiers
1. 25970
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2H4"
1. 1822
/gene="SC2H4.01"
<1. 1822
/note="SC2H4.01"
/note="SC2H4.01, probable ATP/GTP binding protein, partial COS, len: >606 aa; contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/transl_table=1
/product="putative ATP/GTP binding protein"
/protein_id="CAA20595.1"
/db_xref="GI:3559957"
/db_xref="SPTREMBL:O86584"
/translation="1DSTAPKAPVITSLSPYQACADICGEGGCTAGSPFQPMNA DFDIGYRRLITSTSKAKAVTGKSVYSDTTPSLSGQVYVSTAKDVRSMGPAE FSPKAPAGATGTWTFDPAALGSGVTTAKDPAEGRDADLYTAGAGWSTLARGD ADNSMLNDGSDPTONAGYAATSAVAVRDSFTVSAVAYLDTISQTRVYNAFGTG SAFVTLISATYKRWENRTAGVYKDPVYRISGLDAPLKVWTHLAVEPTKDDTN KANDITQLFNGRPOGEVYLDGVSAYQPNVSSGGLGRLVGGANGEHRLGDE VVWVORVLPDEITQOALLLEGVAVDMVAQMDAASGTTGKELSSVPAASRLISA AGAVLDEENNALVIDGSAAGYATGVDNERTSPVSAVQIDSKASKIPGYEALV AGQASAGSSSMALNVRADSVYQKRTIRIVAGDGTQSAEPGGDIAETDITWQ VYGVDAQGEFMTDPAADSAKTEIRGRLHLYGEGDPSFASGTTPOYGAGALAV GSGSRGTGHLHPALRELRYWTGMSADQVRSOVLGV"
188. 211
/gene="SC2H4.01"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
1949. 1953
/note="possible RBS upstream of SC2H4.02"

FEATURES
source

gene
CDS

misc_feature

RBS

gene
CDS

1959. 8510
/gene="SC2H4.02"
1959. 8510
/gene="SC2H4.02"
/note="SC2H4.02, unknown, len: 2183; very limited similarity to proteins of the RBS family e.g. WAPA, BACSU wall-associated protein precursor (2334 aa), fasta scores; opt: 141 z-score: 296.8 E(): 2.7e-09, 23.4% identity in 1237 aa overlap, and RBSA, ECOLI RBSA protein precursor (1377 aa), fasta scores; opt: 135 z-score: 222.4 E(): 3.8e-05, 23.0% identity in 1016 aa overlap. Contains possible membrane anchor around aa 1900. Alternative start at aa 27 would give N-terminal signal sequence"
/codon_start=1
/transl_table=1
/product="hypothetical protein SC2H4.02"
/protein_id="CAA20596.1"
/db_xref="GI:3559958"
/db_xref="SPTREMBL:O86585"
/translation="MRHHPGLVPGAGTGPCHRRRAVPMNRVSAVVGIALPCTT PATADVDPLGRPLSTPRADHVAFKARADAVVRSADQAGAPVYVCS RGTSMPTGSAARLTSAGDAEAPGRPLTAPLAPAKKARADRVYDOKAAR LGKGVVLYKVGPRADGAGARIGIDSARAAVGGDMARKDLYRLPDCTTCPPAPV KTNRGRERLADIAPKASGQTLALAGASADGYKATPLAPSSVTEGGS SCSEFTWSYPLAAPERAPAEKLEISYDGSVDGRTSNNOGTEVGGEDITSFIE RRYGCTDDGHDNRHDLCKYDNASVLYNGKATELVKDDITGOMRLKNDASKVHSG GNDNDNDGEYWTVTGDTGFVEQNTLDGAADVATKSWTVYVEFGDEGKVAL GGFPSRDEKQARWNLDYADETHGNATTYVEDETNVYDLGDDDTISYHGGYLV ERKQVAGALFESGPOASHVSHVYGHDRCLASGCDALTEDPNRDPTVETKCY DDCKPVDYSGSPFRKRLTSVSYTAMDMAATAPSPVDVTLEHLIDPCTGCTG POSTLDEIKTKRKGDTLSIDVAKFSEVPMNRVRSYDSDITSEFRRLTVVETG AOTIVDYLPACVAGQMPKADENKRCFPEVSYGQEPILMFQYVPSVYRTIT PLGSEAVQHTVEYSGGAMWYNDPLTPAEKRWMSGYKTYHYGPGSGRAKT VYVYLRGMDGRVLEGDKTDPDRRAEYSGVYAAVTSODLAGRQSVAYDQ KEVSTVNDPMSKRTATQKSYADEAVYVGVASHARTITSLNPNDRRTTGYC DQGMVPEVDELDGSDVSGDKRCRTTARWADAGLTRVYRTIVGACVYTSALD LPADARPDVYSKATRYDPTTTSATQKPTKCDKMSGRKAGGDDRPIMOTYVLI DITDLGRFVQKNTDVTSTKTEYTOPPAAGPLSTVYVNAANGHRTTYKDALDAVI VTDANGKTESAYDSLRVTSWLPENSRALGKATANYVYGSVSTSLPWSATLNC DSGGRVITYEYDLSLRTVOVASAOGGRIADQLXGRLPVTQAQDINDVDAAPA GKIVQIDGQAPRODVSVDGMRGRTKATVTSYETMAVATVARGDLVILGABEGGS ANAVYDAFGRTVERBDAAGTOPAGDTMTTAYVAFDAADRGKSTIADRSAMVTVYI FGROVSVTPDPRKGTIVYEDALDAVYSDGRGVLLFEYVILKRTGMSAATGAN KLANSEPTLAKGQDTRVYREGGETRAITQKRYTRDPLTKVNNELTLPANDPLA AGVEPARLFTSTGYNLDGTVKQAAAPVAGLSAEVSYTYDGLGVNLAKTGTGLQDA ASPLGDLROMTLATDPTGAKVYLNDYENGRITRSTYTDVDFGFMLOELKYOD DAGNTISVSDATITLGCTGKADHOCFTDGHRLSEAMVETADCSGRTVYAGGAA PYMTSYOYDSDGSLRSKOTENRSGDSDVTTEYGTATGOGPHALSATYTGAEASVYTD EYGMNTERPGVRAOTQIDMNAEGRLAGVSEPAAGKATPATGATVAYYDAGDILLRPTT TIDGELYLYLTGTEVHLKYSNGAKALSAGARTYKAGSVINAVRTSIGVSGTKLITLA GDHGTSGLAINDOTLAFKRWSTPFPGAPROTAGAMPDDGFLGKPADADGTLTQGA ARQYDPTGRFLSDVPLEPDKPNTLNGVAYASNSPYTNSDPSGTSGLGLLGAIGA IIGGVAVIVAGVAAITAVASLGGGGGGGGGTAPASGGMTOPLTKOMTPGATVYE IITKMDLEFPNPSQSLSEMLASMPDMGIVSPRKANWERTSRSLFFPMINGGVPLRE HODFRGGAPFTSILAODETTISGLRSKVGKAKGRTCKTAKENGVFOYVDGPRGSGSW YKNSIGGAADIDINGVFNAGVGTGNQADATLQITSGKARSLINKKEGSDYTLFSAW NGSNRSKTHVYFPRKSNPAFEDTGAAVREDFSWEEKPNINECVNYSWLE"
8536. 8571
/note="hairpin loop with 17/17 bp stem"
8602. 9063
/gene="SC2H4.03"
8602. 9063
/gene="SC2H4.03"
/note="SC2H4.03, possible lipoprotein, len: 153 aa; contains possible N-terminal signal sequence with appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=1
/product="putative lipoprotein"
/protein_id="CAA20597.1"
/db_xref="GI:3559959"
/db_xref="SPTREMBL:O86586"

stem_loop

gene
CDS

```

/translatiion="MULTIPAPAPVRMTAAVLTITLAAATACDGTGRRATVADRTGGA
MOAGGATRLTSDRPTFASGLAGCGGPTGGGMAFFADDDGGLYMSPPAR
RSSEIGLSRREAGSDCECRILTIVVDGDTLCAITSDPETCDLGYRTFRKD"
8647..8679
/gene="SC2H4_03"
/protein_id="P00013" Prokaryotic membrane lipoprotein Lipid
attachment site"
complement(9074..10282)
/gene="SC2H4_04c"
complement(9074..10282)
/gene="SC2H4_04c"
/note="SC2H4_04c, probable aminotransferase, len: 402 aa;
similar to many e.g. AAT_BACST aspartate aminotransferase
(EC 2.6.1.1) (395 aa), fasta scores: opt: 657 z-score:
607.8 E(): 1.3e-26, 33.9% identity in 386 aa overlap.
Contains Pfam match to entry PF00155 aminotran_1,
Aminotransferases class-I, score 136.00, E-value 6.9e-37"
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="CAA20598.1"
/db_xref="GI:3559960"
/db_xref="SPRMEHL:086587"
/translatiion="MERGSKNKSITSGYICRGYPVIEHANALEAGHSYLRANTGNPAL
EGFEPEPITVQDMIRMLPRHAGSTDSRGILSARAAYAOYQALGLEVDVDFLGNGV
SELTSMVAOYALEEGSDYILPADPEPLMTAVYATGLGKAVHTLCEDEAGVCPDLDAE
AKIDRTEAAVYIIPNPNPTAVYAKREIVGIIIDLRARGLVAMLADEIYDILYDDAVH
HSAASIALDPLVYIFGCLISTYVAVAFRSGMVLVYGRPHADYDEGLITLMSRLRCA
NAPAOYALQALAGRSYIIRLTLPGRGLHEQGRVYMERKNEIIPGICSPKRGALYATPP
HIDPVAHVIDDEFEYDILLRRIQVVGCTGGVNMSPDHFILTLPAEDLEAIAIR

```

Query Match	10.18;	Score 91;	DB 2;	Length 25970;
Best Local Similarity	47.28;	Pred. No. 9.9e-12;		
Matches 422; Conservative	0;	Mismatches 455;	Indels 18;	Gaps 4;

OY	13	ggaattttbtaagagggtaacacactatataagttgcgaattgaaacccaatttgtaattgga	72
Db	24459	GGCATTCACGTAAGCAACGAGCGTCTCTCGAGGGCAACGGAGGCTGTCTCATAGGCG	24518
OY	73	ggagataacacatgggcacgcacgtgtatataaagacacagacactactgcgaattgaagagatt	132
Db	24519	GGGGTCAACACAGCGCTACACCTGGTATGCCAGC---GCACGGGCTCCATGCGGACATTC	24575
OY	133	gcaaataccggtgtcctaatacgggtccggatctgtgtatctatgtaabggggaacaagaa	192
Db	24576	GCGGCCAAGGGGACCAACCGTCCGCGTGGTGCATCAGCAGTGGCGCGGTGACGAAG	24635
OY	193	gatgacacacataacagtaagaacacttctcttaaggggaagataatcattggttgc	252
Db	24636	ACGACCGGCTCCGAGGTCTCGCGCTCATCGGCCACGTGCACAGGCCACAAAGGTCACTTGC	24695
OY	253	gttcttgaagttcatgatcgtctacccggttatgatt-----ccattgcttcgtcaat	303
Db	24696	GTCTTTGAGGTGGCAAGCAACACCGGCTACGGCGGAGACGGCGCCGACCAAGCTTGGAC	24755
OY	304	cggtcgttgtattatgtgattgaaatgaaatgctttaaattgaaagaaagatacgtc	363
Db	24756	CAGCGCGCGCACTACTGGGTGAGTGAAGACGGCGTGAAGGACGAGAGACTACATCGTC	24815
OY	364	attataatatgtcgaaatgaatgattggttctgtgtggaagggatgagcttggctgtacgg	423
Db	24816	GTCCGCAACATCTGGCAACGACCTTTCGGCAACACCACTACACCGCGGTGGACCCACCG	24875
OY	424	tataaacaagaacatcccgcatgtgtgaacgcggtctaaacacataccttgatgttagat	483
Db	24876	ACCAAGTGGCGCATGGGCAAGTCCGCGGCGCGGCGGCTCGACACGCGCTGATGTTCAAC	24935
OY	484	gctgcggggttggagc---aatttcacaatcgaattcagatbataltgaaagaaagtttt	540
Db	24936	GCGCCCACTGCGGCGCAGAGACTGTCCGGAAGATGGGTCCACAGCGCGCTTCGCTTC	24995
OY	541	aatgttcgacctcaacgaaatacatgttttggattcatatgatagaatatgcagttgct	600

Accession	Sequence	Position
D82499.6	GCCCTCCAGACCCGACCGCAACCCCTTCTCCGCGCCAGCAATGAG---GGGTCTACGAC	25052
QY601	aatgatgcgaagtgcgtaactaataatggacggagttcctaactgaactctgaaatagtc	660
D825053	ACCGCCGCCGAGAGTCCGGGACTACCTGAAACGCCCTTCGTGGCAGCGGACTGCCATCTCTC	25112
QY661	attgttgaaatttggacaacgcctcatacaaatgtgtgacgtcgtatgaaacagattatgac	720
D825113	GTCCGTGAATTCCGGCGACACACAGTACCGGCAACCCGACGAGAGCCATCATCTGGCG	25172
QY721	tattctgacaagaagggtgtgtgtgtgtgtgtgtgtatgtgaagggaagggcccaaa	780
D825173	ACGGCCCACTCCCTCGGGGTGGCTACCTCCGGCTGTGCTGAGCGGGAACGCGCGCGC	25232
QY791	ttggaggtatttagaccccttcgaattggtgtgtgaaataaacctbacaagcttgggaaat	840
D825233	GTCCGAGTACCTCGACATGGTCAACGCGTTCGACCCGAACTGCTGACAGCATGTGGGGCAAC	25292
QY841	acaataaggaatgtgcatatgtgtttaaagaaataatctgaaatgaacggttt	895
D825293	CGCATCTTCTACGCGACGACGGCATCGCCCGACGAGTCCAGAGAGCGGCCACCGCTGT	25347

RESULT	8
STREPMANASE	
LOCUS	STREPMANASE 1881 bp DNA BCT 16-JUN-1999
DEFINITION	Streptomyces lividans mannanase (mana) gene, complete cds.
ACCESSION	M92297
VERSION	M92297.2 .GI:4579683
KEYWORDS	.
SOURCE	Streptomyces lividans. Streptomyces lividans. Bacteria; Filicutes; Actinobacterlia; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 1881) Arcand,N., Klupeff,D., Paradis,F.M., Morosoli,R. and Shareck,F. Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence of the mana gene and characterization of the enzyme Biochem. J. 290, 857-863 (1995)
JOURNAL	Biochem. J. 290, 857-863 (1995)
MEDLINE	93207541
REFERENCE	2 (bases 1 to 1881) Shareck,F. Direct Submission Submitted (28-APR-1992) Centre de recherche en microbiologie appliquee, Institut Armand-Frappier, 531 boul. des Prairies, Laval, Quebec H7N 4Z3, Canada 3 (bases 1 to 1881) Shareck,F. Direct Submission Submitted (08-APR-1999) Centre de recherche en microbiologie appliquee, Institut Armand-Frappier, 531 boul. des Prairies, Laval, Quebec H7N 4Z3, Canada Sequence update by submltler On Apr 8, 1999 this sequence version replaced gi:153193.
AUTHORS	
TITLE	
JOURNL	
COMMENT	
FEATURES	
REMARK	
CDS	location/Qualifiers 1..1881 /organism="Streptomyces lividans" /strain="1326" /db_xref="taxon:1916" 217..1316 /gene="mana" 217..221 /gene="mana" /note="putative" 226..1377 /gene="mana" /EC_number="3.2.1.78" /codon_start=1 /transl_table=11 /product="mannanase" /protein_id="AAA26710.2" /db_xref="GI:4579684" /translation="MRNAARSLITTAGCAFAVILGLPALGSPAGRAEAAAGAIGHVSN"


```

QY 73 gggatataacatgagcagcagatgataaagaccagcgaactactcgaattgaaggatt 132
DB 61 GGGGTCAGCCACCCCACTGCT---ACCCCAACACACCCAGCCGTCGCCGACATC 117
QY 133 gcaaataccggtgctgaataacggtccggtatgtgtatctcgtatggtgggagacaataa 192
DB 118 AAGTGGCAGCGGGCCCAACCCGTCGGGTGTGCTGACGACGGGTTCGGGTGGACCAAG 177
QY 193 gatgacatccatacagtaagaacctatctctttagcggaaagataatcatttggttgt 252
DB 178 AACGGGCTTTCAGAGTGGCCCAACGTCATCTCCCTGTGCAAGACGAGACGCCCTATCTGC 237
QY 253 gttcttgaaatcagatcagatcagcgggtatgattccatt-----gcttcgtcaat 303
DB 238 ATGCTGGAGGTGACGACACCCGCTACGGTGTAGACAGCGGGGCTTCACAGCTCGAC 297
QY 304 cgtgctgttattgattgattgaatagaagtgcttattattgaaagaaataaccctc 363
DB 298 CAGGCGCTGACTACTGATCGAGCTGAGAGCGCTGCTCAGGGCGAGAGGACTATGTC 357
QY 364 attataatattgcaaatgaatgattgtt-----tcgtggaaaggaggtgcttggt 417
DB 358 CTAATCAACATCGGAGACGAGCCCTACGGCAACGACTCCGCCAGCGTCCGCTGGGGG 417
QY 418 gaagggtataaagaagaactcccgagattggttaacgcccgtcttaaacataccattgatt 477
DB 418 TGGGACACCTCCGCGCCATCCAGCGGGCTGCGCGCGCGGATTCGAGCACACCTCTGTC 477
QY 478 gtagatgtcgggggtgtgggac---aattcacaatgattcatgattatgaaagaa 534
DB 478 GTGGAGCCCCCACTGGGGCCAGGACTGAGACGAMACCATATGGGAAACAGCCGACCA 537
QY 535 gttttaaagtcgacctaacgcaataacatgattcgtattcatatgatatgaata 594
DB 538 GTGTACGCCAGGAGACCCCAACCGGCAACCGCTTCTGATCACAATGATACG---GCGTC 594
QY 595 ggtgtgaatgatacgaagtcgttactaaatattgaccggaattcctaacaagcctcgca 654
DB 595 TACTCCCAAGGCTCCACGATCAGACTACCTGAGACCACTTGTCAACGCGGCGCTCGCG 654
QY 655 ttagcattgtggaatttggacacgctatacaaatggtgaagtgatgaagaagatt 714
DB 655 CTATCATCTGGGAGTTCGGCCACGACCACTCCGAGCGCAACCCGACAGGACGACGATC 714
QY 715 atgaactatctgaacaagaagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 774
DB 715 ATGGCCGAGGCGGAGCGGCTCAAGCTGGGCTCATGCGCTGCTGTGGAGTGGCAACGCG 774
QY 775 ccagaatggagattattagaccttgcgaatgattgggtgtgaaataacctacagcttg 834
DB 775 GCGGGGGTTCAGTACCTCGACATGTGTACAACTTGCAGCGGACGACAACTGAGCCGCTG 834
QY 835 gg 836
DB 835 GG 836

```

```

RESULT 10
SCF73/c 14922 bp DNA BCT 04-OCT-1999
LOCUS Streptomyces coelicolor cosmid F73.
DEFINITION AL121746.1 GI:6013076
VERSION AL121746.1 GI:6013076
KEYWORDS catalase/peroxidase; cpeB; Fe regulatory protein; furs; manA;
membrane-bound oxidoreductase; Merr-family transcriptional
regulator; polyphenyl synthetase; secreted beta-mannosidase.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 14922)
AUTHORS Redenbach, M., Kleser, H.M., Denapalte, D., Elchner, A., Cullum, J.,

```

TITLE Kinashi, H. and Hopwood, D.A.
 JOURNAL A set of ordered cosmids and a detailed genetic and physical map
 MEDLINE for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 97000351
 REFERENCE 2 (bases 1 to 14922)
 AUTHORS Seeger, K.J. and Harris, D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 14922)
 AUTHORS Parkhill, J., Barrell, B.G. and Randle, M.A.
 JOURNAL Direct Submission
 TITLE Submitted (04-OCT-1999) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

by the BBSRC and Beowulf Genomics
 details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh1.go.jp/>.
 CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
 Cosmid F73 overlaps with cosmid F11 on the AseI-F genomic restriction fragment.
 Location/Qualifiers
 1..14922
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid F73"
 1..122
 /note="nominal overlap with Str11"
 complement(1..983)
 /gene="manA"
 complement(1..983)
 /gene="manA"
 /note="SCF73.01c, manA, probable secreted beta-mannosidase, len: 327 aa; similar to many e.g. MANA, STR11 P51529 mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189 z-score: 1332.7 E(-): 0, 61.1% identity in 288 aa overlap. Contains N-terminal signal sequence, and Pfam match to entry PF00150 cellulase, Cellulase (glycosyl hydrolase family 5). Also similar to SC2H4.16 (62.2% identity in 288 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="putative secreted beta-mannosidase"

TITLE
 JOURNAL
 MEDLINE
 94150718
 2 (bases 1 to 35049)
 Latreille, P. and Gattung, S.
 The sequence of C. elegans cosmid K03E6
 Unpublished (1998)
 3 (bases 1 to 35049)
 Waterston, R.
 TITLE
 JOURNAL
 4 (bases 1 to 35049)
 Robert Waterston
 Submitted (17-APR-1996)
 Waterston, R.
 TITLE
 JOURNAL
 Submitted (21-APR-1997)
 5 (bases 1 to 35049)
 Waterston, R.
 TITLE
 JOURNAL
 Submitted (28-OCT-1998)
 Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Oct 28, 1998 this sequence version replaced g1:1280123.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RO, England
 e-mail: rtw@emslab.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is C04D1, 2700 bp overlap; 3' clone is F49E7, 200 bp
 overlap. Actual start of this clone is at base position 701 of
 CELK03E6; actual end is at 35049 of CELK03E6

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program GeneFinder (P. Green and L. Hillier, ms in preparation).
 Location/Qualifiers

source
 1. 35049
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /clone="K03E6"
 /chromosome="X"
 complement(3691..7748)
 /gene="unc-1"
 complement(join(3691..3766,3931..4145,4822..5095,
 5154..5278,6432..6545,7296..7326,7735..7748))
 /gene="unc-1"
 /note="K03E6.5"
 /codon_start=1
 /product="Erythrocyte band 7 intergal membrane protein"

/protein_id="AAC69044.1"
 /db_xref="GI:1280125"
 /translation="MEYSHQIPKSCATIDVPDPYETIGTIFGALQALSMILLIIVT
 PFSMCVCLIKYKEKREYVIFRIGLIVGARGGPMIIFICIDYRQIDRVVAVY
 FOELISDVSVDVAVYFETSDIPISVNNVDIAISTLQTLRMAAEARAEARF
 MLTEREALIAOLCETIIDEGETHEGVCYKEREVVDIIRPOOLTRMAAEARAEARF
 VVAEGEOKSRALKEADYIQANPVALQRLHQAALNSIAEHNSITVFPEVMEGA
 PMKDDQ"
 complement(12649..14893)
 /gene="K03E6.4"
 complement(join(12649..12750,12801..12928,13397..13527,
 13570..13666,14217..14391,14837..14893))
 /gene="K03E6.4"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC69043.1"
 /db_xref="GI:1280126"
 /translation="MEKADSVITLHGNVTEPEPESNYCVKISTVANGDKTINCE
 TVLSISFSGADNQIFIDLPFCGLKSSPFPMRPRLVYLCFSKDSGRKCVGYGML
 PIPPEKHCYKRVHCFLEPDSVVOQLAKRVNAEFVPMPLPANADKRYACRTSTK
 GYVDELTNVISIKCDTAKFTSTETIKSSNSHATNVSTTIADLGEAVGNEETFN
 AOEEDDESKR"
 20796..24499
 /gene="11m-6"
 join(20796..20881,21727..21835,22141..22262,23469..23598,
 23885..24151,24209..24499)
 /gene="11m-6"
 /note="K03E6.1"
 /codon_start=1
 /product="similar to other homeobox domains; contains two
 LIM domains"
 /protein_id="AAC69042.1"
 /db_xref="GI:1280127"
 /translation="MSLLISATSTTEDEKLGCGGCLIKRFFLTIRIKSKQOMA
 HLLVDFILITIVLDSKSNVMEDESHESCLSCCGLISFCKFCFSGHNYCEH
 DHMLYGRRCRCMTLLPDIYHVRHYHQCSCGSCORPMNGDEYVHDEY
 FCGNDOSINFGTSEHYLTFTFAISVNNVDIAISTLQTLRMAAEARAEARF
 AORROKTEKFEKSSKPSKVRQQLANEGLSIRYQVWFONRAIKIKINKDSDSG
 DTRKHGSGRSTEDIRSDDEESYISLKRIGIDIGELMLIKLSVFOKRVKRPV"
 complement(25406..26485)
 /gene="K03E6.6"
 complement(join(25406..25503,25556..25679,25726..25773,
 26375..26485))
 /gene="K03E6.6"
 /note="similar to Profilin; K03E6.6"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC69045.1"
 /db_xref="GI:1280128"
 /translation="MSWSDIINNINIGSGNVSKAALIGFDGAVMAKSDNPNISVEAV
 AKKATSIDALGTLGRGQFVFLVNDNDKITIGKGGSGFYIKTIQAVIISYE
 KGIQPKCSKTGTGALDYFRSIRY"
 complement(28061..33962)
 /gene="K03E6.7"
 complement(join(28061..28224,28960..29077,30844..30997,
 31048..31205,31305..31629,32214..32247,33211..33354,
 33555..33654,33708..33861,33914..33962))
 /gene="K03E6.7"
 /note="similarity to C. elegans protein C01C10.4"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC69046.1"
 /db_xref="GI:1280129"
 /translation="MDETATSEVETFEVSDPTTQFEEDGHPLETIRHLNMIHELE
 KINISTDIVNKMEVOLDARADRFETOVMSKLELSKQVSOIAKAPFELIKE
 RSLREESQKAERFERATSIILGIAKQVSLTQESLSRQSVLPECLEVNHIQRY?
 VEERFAASLSAKSKAMHLAERTKRAEKNNRYAIKRSRYPERKLEETITLNGX
 ARTICLEAVROKKNDYTSLNLEISIRIHERBSTGLSEAVSDDQDQSDKSS
 ESUPNPYPYATAPPTDYKTIIDDDDSIYLNMIKTEQEEGERKRSLSGSGVIL
 LMQQLIGNNSTEKHNITPPRRGEEDISYHVRPVGLSGDSNSESLSASFNISGSL
 TVSKMMSHSELKRCDAIKIAISVIAAKKFFONKQPLQVADKFRPLRPYKXK
 INSTLISIDTFLANQIVFEISMAVGERGYDFPFLSLECFRVSGETCR"

DEFINITION Helicoverpa zea nuclear polyhedrosis virus, partial sequence.
 ACCESSION U67264.1 Z12117.1 U03418
 VERSION U67264.1 GI:2078305
 KEYWORDS
 SOURCE ORGANISM
 Helicoverpa zea nuclear polyhedrosis virus.
 Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 Nucleopolyhedrovirus.
 1 (bases 335 to 1075)
 Cowan, P., Bulach, D., Goode, K., Robertson, A. and Tribe, D.E.
 Nucleotide sequence of the polyhedrin gene region of Helicoverpa
 zea single nucleocapsid nuclear polyhedrosis virus: placement of
 the virus in lepidopteran nuclear polyhedrosis virus group II
 J. Gen. Virol. 75 (Pt 11), 3211-3216 (1994)
 95053907
 2 (bases 1 to 14235)
 Le, T.H., Wu, T., Robertson, A., Bulach, D., Cowan, P., Goode, K. and
 Tribe, D.
 polh-1el genomic region of HzSNPV
 Virus Res. (1997) In press
 3 (bases 1 to 14235)
 Le, T.H., Wu, T., Robertson, A., Bulach, D., Cowan, P., Goode, K. and
 Tribe, D.
 Direct Submission
 Submitted (20-AUG-1996) Microbiology, U. of Melbourne, Grattan
 Street, Parkville, VIC 3052, Australia
 On May 14, 1997 this sequence version replaced gi:60442 gi:454319.
 COMMENT
 FEATURES
 Source
 1. 14235
 /organism="Helicoverpa zea nuclear polyhedrosis virus"
 /strain="Elcar"
 /db_xref="taxon:10468"
 335..1075
 /codon_start=1
 /product="polyhedrin"
 /protein_id="AAB54089.1"
 /db_xref="GI:2078306"
 /translation="MYTRYSYSPYGLTYVYDNNKFKNGAVIKNARKKHLSEHEE
 ERNLSLDKYLVAEDPFLGPGKNOKFLFEIRISVGRDPMKLVYVNGSGRPIETWRE
 FMEDSEPIVDQELMDVFLSYNNRPKPCNCRRLAOLHAKCDPDYIPHEVIRAYEES
 YVGSNNERISLAKRYGGCPVNNIHAETNSFEDITVNNETKPIYVGDSEAE
 EEILEVSLIFKIEFADAPLYTGPAV"
 complement(1072..2310)
 /note="AcMNPV ORF8/ORF1629 homolog"
 /codon_start=1
 /protein_id="AAB54090.1"
 /db_xref="GI:2078307"
 /translation="WVQLQSYQOYLAKENNTNNVDFLLRMISAPSLKRTVSAVQ
 SNSDRQLNVDICQLKLAQEIYNNVDLIDIGRSIIPILSQTMSVPYVPTTIK
 PTVVSTEDIMQKTSPIIIDSSESSQTLAPPIPPPIPPQTINSSITIPPE
 PPPPLPILEDIVIPSSKDRDYTSRQVLRKPRELMEQIQGIKLKVSAPDG
 SIYVTVAAASPTAKILORIRAVQSPVSESGWGTDEQQOQASSELKQYRSYX
 NITLDSWIKNSYSTRAQDTLISIKRQNLORLSNATQGISAKLQIFIDNLIQDTY
 DNPLDKSDNDELQSDKNLQFIMAVBDLFFKQYDALANVMASALGTPTNKLQEL
 RSNLDKIMSKTWTSTESQV"
 2325..3128
 /gene="pk1"
 /gene="pk1"
 2325..3128
 /codon_start=1
 /protein_id="AAB54091.1"
 /db_xref="GI:2078308"
 /translation="MDRRFVKEINQFAETIKIONNVRVLDGKSGKGVIRHEPTGKLF
 VKRSVAIKYTEIEPLVHOLMKNRVFIKLYSLTLKSOILIDYVAGDDLDDELAK
 HKRYSEATRSYNGQLTEALNAHLSYIIINDLKEVLYVRHQIILCQGLACKLYN
 TSSCRDGTKEIMSEPKLRQNYDVHVMALGITLYELLGHHPRKYSNNEDFDLD
 VLQQRQOKKLAKKNFLSSDAOKFLEAMIMNINRICTYETVLIKHSFLS"
 3105..3164
 /note="dyad element"
 3178..3203
 /db_type=direct
 repeat_region
 3224..3249
 /db_type=direct
 repeat_region
 /rpl_type=direct

CDS
 complement(3384..5522)
 /codon_start=1
 /product="HoAR"
 /protein_id="AAB54092.1"
 /db_xref="GI:2078309"
 /translation="MPSTNKRVRYSIDSSSSSEAVSVOLNNHKTLEISNISEORIY
 VYVILSNKRRLGTYVCKTRFRPMSGVATANKCHVYNSKLEFDLDYHSTWGSCLDSGE
 FENPVTLITKAMERKINRYAVQDLYAVLANLSDSSISYATQKFIKNNKANNKEAI
 VQCLNMTKTDSSASVDNGISLMDIRHQLFLNKECPAFVNEHDLRIEIQKTKT
 KHAQVYKQHRDLVITKCSYCNHSDITLTPQCMHRCSCSLRSIQINCMATKRNK
 TSRNSDGDSDSDSDSEYDEVVANNVNNNDNNDNSDSDSDNSNCDDATIHNEPT
 AEINTNTVNDIDETNNDASVNNNTVNDQDDDDDDVNDVNDVNDVNDVNDVNDVNDV
 NVNSDNTVTNTDNTVNDVNCNTNANNVIANVDTNANVDTNANVDTNANVDTNANVDTNANV
 NDVQHVYDNDVNNVNNVQTSMDNINSNGVNNNSKNDKNSVDDDDDISNLEPTYN
 ITVLIRYVDNHCNDSNDVFPSSNSNNNSNNNSNNNSNNNSNNNSNNNSNNNSNNNSNNNS
 SQNLNLSQLENEVRNLEKSLSTVLPVAMSTEEELSDIKELAKSTNDKNGKYEYH
 ITVTEKTSKNDATNTDNNDAAPFDIAKELGDDFNECEPNDYINEYHECVPKQ
 EPVMSDVECLANTPLSPASSMTIVNRK"
 complement(5753..5777)
 /note="conserved DNA repeat element; similar to AcMNPV
 len/pe38 promoter element"
 len/pe38 promoter element
 complement(5815..5841)
 /note="conserved DNA repeat element; similar to AcMNPV
 len/pe38 promoter element"
 complement(5886..5910)
 /note="conserved DNA repeat element; similar to AcMNPV
 len/pe38 promoter element"
 6442..6924
 /note="ORF480"
 /codon_start=1
 /protein_id="AAB54093.1"
 /db_xref="GI:2078310"
 /translation="MPRRSVRSLSDDKRYKFEKAVYPPDVSRKYLPPKLDKTRFE
 LFEDLMNVRKRKHFDNDSDSDNDNDNISQVADILSHIRYHREQAKQDTISVIFZ
 LENKSFLLAMILEPLIDQYNSDFLFIKIQNSKYYNHSFLSDIDGSDYRDLDPYFI
 6664..6672
 /note="putative splice donor site"
 7279..7602
 /note="ORF321"
 /codon_start=1
 /protein_id="AAB54094.1"
 /db_xref="GI:2078311"
 /translation="MSGTLKRLIYDDISDSDQAKLFRYNSQMPPASQOMNTAVDY
 IDVEYIKCKERKKNYSDDVITNARAQYNNKLAFLVLDDEYKKQYNNNDKQSVLYIK
 EISNL"
 7680..8135
 /note="AcMNPV ORF1 homolog"
 /codon_start=1
 /protein_id="AAB54095.1"
 /db_xref="GI:2078312"
 /translation="NMFSDENVYVPKRYFFTLVLYNKLKLEFDYDCKCAKTLQGL
 QALNTLQLQASASCKHIAHINQSOVLVFFLENPYEENICDITDNRHRHNPVCC
 GYKICMICYANLMTKYSTVFPCTVCYKTSFSSSSVSFKQVYTADTTNT"
 8152..9558
 /note="AcMNPV ORF2/p49 homolog"
 /codon_start=1
 /protein_id="AAB54096.1"
 /db_xref="GI:2078313"
 /translation="NNIDENKVALDENNYKYLFLASVFNLDATGLTSSKPIREYL
 YNNFNNDIASLGLDYLDLIGLVNVLIDBDVNNMKYIKDPREFGCTKKNVLEED
 ORYIKDPDPVYATNFVKNPSERFLVYNNVSSVYDKRNFVNDKNGVLIQNTCYM
 FQDAYVDKGCVRCEYPRKLEDESWPPRYALVYLGDMACAFATNINISPGNGTILNF
 YKGLPEFRNKILINSKRTTKKPNHLNFEQKQEFDSKAYVFIQRDVYDKAAPD
 DLDLNLNEHTYTSYVRYKQEMDEDEEGNYSEIVDRYAVDKQKLSIKIDEYTM
 FPLIRANDSYIIRPDLIQIGTLNAFVYKSKLEPAIILANNSLFGSTLLLEDRKLI
 PYQYQPPYRLANDEYVVDKQGLVLYKTFTANTIPAVYLLINDYESSSEIKTLRDLK
 PAVQNTLKLILIAAPSK"
 9569..9814
 /note="AcMNPV ODV E18 homolog"
 /codon_start=1
 /protein_id="AAB54097.1"
 /db_xref="GI:2078314"

```

/translation="MDDLGGTTTGGAGFNNPMLNPSMLTLLALVLIILLIMLFOS
SSPGSKGDTNAFAFQNPFLNATMNNPPVNTPORTMM"
9830..10684
/contig="ACMPV ODV EC27 homolog"
/codon_start=1
/protein_id="AAB54098.1"
/db_xref="GI:2078315"
/translation="MKKRCOSNRIETETELINADERLHKVDLADFAKMLNLESTY
DNLQIKMLAKYAMLANLETOPILAFPRKNATREYVSYVPSLGVNRRVPMIN
HNSKMEFIVENRNASIPGEPLFCQNDVDVCIIDRPSILOMLSDPLDVAVN
MHKSNKMYIAKTRFCAPKRHRREPPLEINTEVDVQMLLPIHSHALAY
ILKNYGVYDYSRSLSDHLFNSKSRPTLMNFSNALLSKFSEIYDYSINTKNTKN
LGILTYND"
complement(11095..11646)
/contig="ACMPV ORF5/EP23 homolog"
/codon_start=1
/protein_id="AAB54099.1"
/db_xref="GI:2078316"
/translation="NINLYQPLSHIDNLIYFTIYDARFVYVFNKANDVINDON
PNTYTRKLSGSGRANYSNMSTTESGEMIIDVYIIMFRLPFAKSLIDBEC
FSPVLLADYNSTETWVLSVRRSNMAKCYERDALFTDSAGELTSKEHVI
IKGNVPAFCVLCOSPROQVVT"
11688..11637
/gene="tel"
11688..11637
/gene="tel"
/codon_start=1
/product="immediate early 1"
/protein_id="AAB54100.1"
/db_xref="GI:2078317"
/translation="MANRITPLRDVGNVNTINYPPOSQSCNNDSDSYMNNND
VDVKKLTVENASKNRYEMASAFASITPTSSKSPRNNHLPFGELYMSPRMS
PQRTPRSESENVTESLPESLSLQVTVSLKRSGLYGNLQNLKNTETMDPIE
SDSSSELTTPKPKRSNTERKIAVGERSKREKATPLNMPGVANNKQDLMDAP
NRKYOVLKPOHPDPDPNCKWPREFEDYRNMPLVNMPFRRKRRDFI
QOKNYHMEIVKEQENINSISIEHLYLVNTVASIYESSYVNDLVHYVTEFRY
FMISRLTLKLNHLPESQFPMYHODRSKCHNEIKQVNMELMDLVNMY
OTELYSLSATGPKRKYLTLSVMEHIDHLYLPITLSQESLEIDIQRYVASVL
YVQNVLSKDVQKQIENFMNDVINYTVALKELSKNEINAVKEOSDFETIK
YGVNRLFKESIRTNALIKRETHAGLIDNTLEAQNNDTSSFLINTKDERI
TIIRKGPLFTLITSLIIDLIAMDLEIKYKKTTHVFNLSTNRKEMNKNHGMKILS
FYTSNLMDELKEFAVNNFNCSD"
complement(13706..14235)
/contig="ACMPV ORF148 homolog"
/codon_start=3
/protein_id="AAB54101.1"
/db_xref="GI:2078318"
/translation="ALRRGGSYYHIGLNGEYQESCLLRRTCYLDVNNIDVNC
SDPLIDNINATOSYGHYNAEYVTCRSDPNADPLSLQYVDSPLATGHTISIEP
YDPGDLIGDLIGLIGEGILNKSDDKSSSTSPKLPDIYVVLGIVLIIITGIYVIK
RMLMQPPPPNRYNR"
BASE COUNT 4624 a 2618 c 2541 g 4452 t
ORIGIN
Query Match 4.9%; Score 44; DB 73; Length 14235;
Best Local Similarity 47.2%; Pred. No. 1.8; Mismatches 150; Indels 0; Gaps 0;
Matches 134; Conservative 0;

```

```

OY 338 cttaattggaaggaagataccgcatattataattgagat 381
DB 4315 CTACTAATGATATGATCTAATGATGATGATGAT 4272

RESULT 15
LOCUS HS1018E9/C
DEFINITION Human DNA sequence from clone RP5-1018E9 on chromosome
ACCESSION 209413.31-13.33, complete sequence.
VERSION AL035455
KEYWORDS AL035455.30 GI:9581749
SOURCE HTG.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 148667)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Jul 28, 2000 this sequence version replaced gi:6065863.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone configs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/RGP/Chr20
RP5-1018E9 is from the library RP1-5 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-1018E9 The true
left end of clone RP4-68519 is at 127330 in this sequence. The true
right end of clone RP11-10939 is at 61512 in this sequence.

FEATURES
Source
1..148667
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.31-13.33"
/clone_11b="RP1-5"
/clone="RP5-1018E9"
BASE COUNT 40526 a 31692 c 32055 g 44394 t
ORIGIN
Query Match 4.8%; Score 42.8; DB 38; Length 148667;
Best Local Similarity 56.3%; Pred. No. 3.1; Mismatches 62; Indels 0; Gaps 0;
Matches 80; Conservative 0;

```

Oy 638 ttaatcaagacctcgcatatgtagtggaattggacaccggtcatcaacaatggtgacg 697
Db 19456 CCAATCAAGATCTCAGCTGGCTTTTGAAGAACTGGGCACTGTGATTCTAAATTATTA 19397
Oy 698 tcgatgaagcaacgattatgag 719
Db 19396 TAGAAATGCAAAAGACCCAGAG 19375

Search completed: December 20, 2000, 03:17:03
Job time: 13068 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:58:32 ; Search time 168.09 Seconds

(without alignments)
2011.401 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_990

Perfect score: 900
Sequence: 1 aatgcacatccgcgatttca.....gattacacacccgttttaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

N_Geneseq_36:*

- 1: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	900	100.0	1470	21	DNA encoding a Bac
2	895.4	99.5	1438	21	DNA encoding a man
3	468.2	52.0	1407	21	Bacillus agaradher
4	468.2	52.0	1407	21	DNA encoding a Bac
5	468.2	52.0	1482	21	Bacillus agaradher
6	468.2	52.0	1482	21	DNA encoding a Bac
7	408.4	45.4	1107	21	DNA encoding a Bac
8	405	45.0	995	21	DNA encoding a Bac
9	304.4	33.8	960	21	DNA encoding a Bac
10	157.8	18.6	564	21	DNA encoding a Bac
11	104.4	11.6	397	21	DNA encoding a par
12	78.2	8.7	915	21	DNA encoding a par

13	35	3.9	2040	17	T29774
14	35	3.9	2100	10	M90712
15	35	3.9	2901	12	Q14810
16	35	3.9	3558	13	Q22596
17	34.6	3.8	580073	18	T58840
18	33.6	3.7	627	20	Z08233
19	33.6	3.7	1395	20	Z08233
20	33.6	3.7	1680	20	Z08234
21	33.6	3.7	1763	15	Q55316
22	33.4	3.7	2408	12	V17621
23	33.4	3.7	2802	19	Q14809
24	33.2	3.7	633	18	V75647
25	33.2	3.7	1800	20	Z25012
26	32.6	3.6	138169	21	A34791
27	32.6	3.6	141589	21	A35005
28	32.6	3.6	141589	21	A35030
29	32.4	3.6	1300	20	V82078
30	32.4	3.6	2132	20	V82077
31	32.4	3.6	2841	18	V74488
32	32.4	3.6	3006	21	Z29484
33	32.4	3.6	22080	20	X06751
34	32.2	3.6	1021	16	T04613
35	32.2	3.6	1026	17	T13340
36	32.2	3.6	1399	20	X20017
37	32.2	3.6	1500	20	X20016
38	32.2	3.6	3111	16	O87260
39	32.2	3.6	4956	20	X13728
40	32.2	3.6	1448	18	H14454
41	32.2	3.6	4359	18	T69969
42	32.2	3.6	4365	14	Q52440
43	32.2	3.6	4365	13	Q25536
44	32.2	3.6	4365	13	Q25534
45	32.2	3.6	4365	14	Q52443

ALIGNMENTS

Bacillus thuringie
cryd gene. Bacill
Btm Pg14 72kDa Cry
Coding region of p
Mycoplasma genital
Lactobacillus brev
Coding sequence of
Lactobacillus brev
Bacillus sp. V230
Btm Pg14 72kDa Cry
Staphylococcus aut
Lactobacillus reut
Human adenosine re
Human adenosine re
DNA encoding a par
DNA encoding a gid
Staphylococcus aut
DNA encoding human
Salmonella enteric
5' flanking region
Listeria phage lys
Enterococcus faec
Enterococcus faec
Flea sodium pump 2
Enterococcus faec
H. pylori GHP0 933
FIPV spike (S) gen
Feline infectious
S gene of TS FIPV.
S gene of wild typ
Feline infectious

RESULT	1
ID	245335 standard; DNA; 1470 BP.
XX	245335;
AC	
XX	
DT	27-MAR-2000 (first entry)
XX	
DE	DNA encoding a Bacillus mannanase enzyme.
XX	
KW	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW	endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW	galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW	printing paste; plant material degradation; recycled waste paper;
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;
KW	mannan-containing food; coffee extract; cleaning composition;
KW	machine washing; hard surface cleaner; dishwashing; oral; dental;
KW	contact lens; body care composition; fabric softener; oil well drilling;
KW	subterranean formation fracture; ss.
XX	
OS	Bacillus sp.
XX	
EH	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1470
FT	/*tag- a
FT	/product- "mannanase"
FT	/transl_except- (pos: 346..348, aa: Pro)
FT	/note- "no termination codon given"
FT	1..90
FT	sig_peptide
FT	/*tag- b
FT	91..1470
FT	mat_peptide
FT	/*tag- c
FT	/*note- "specifically claimed in claim 4; nucleotides
FT	91-990 encode the catalytically active core"

```

XX  MO9964619-AZ.
XX  16-DEC-1999.
XX  10-JUN-1999; 99WO-DK00314.
XX  10-JUN-1998; 98US-0111256.
PR  20-OCT-1998; 98DK-0001340.
PR  20-OCT-1998; 98DK-0001341.
PR  28-OCT-1998; 98US-0105970.
PR  28-OCT-1998; 98US-0106054.
PR  23-DEC-1998; 98DK-0001725.
PR  05-MAR-1999; 99DK-0000306.
PR  05-MAR-1999; 99DK-0000307.
PR  05-MAR-1999; 99DK-0000308.
PR  05-MAR-1999; 99DK-0000309.
PR  09-MAR-1999; 99US-0123543.
PR  10-MAR-1999; 99US-0123623.
PR  10-MAR-1999; 99US-0123641.
PR  11-MAR-1999; 99US-0123642.
XX  (NOVO ) NOVO-NORDISK AS.
XX  Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX  WPI: 2000-105891/09.
XX  P-PSDB; Y54122.
XX  New mannanases for treatment of textiles, plant material and coffee
XX  extract, and in cleaning compositions
XX  Claim 4; Page 208; 242pp; English.
XX  The present sequence encodes a mannanase (also known as mannan
XX  endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
XX  of Bacillus sp. 1633. The mannanase hydrolyses galactomannans.
XX  Specifically, mannanases hydrolyse 1,4'-beta-D-mannosidic linkages in
XX  mannans, galactomannans, glucomannans, and galactoglucomannans. The
XX  mannanase protein, or preparations containing it, are used to improve
XX  properties of cellulosic or synthetic fibres, yarn or (non)woven
XX  fabrics (removal of mannan-based sizes or printing pastes). They are
XX  also used to degrade or modify plant materials (particularly recycled
XX  waste paper, paper making pulps, or material containing guar or locust
XX  bean gums (thickeners), or to reduce viscosity of mannan-containing
XX  foods or feeds). The mannanases are also used to process coffee
XX  extracts (to inhibit gel formation); in cleaning compositions (for
XX  machine washing of fabrics, as hard-surface cleaners, for hand or
XX  machine dishwashing, also in oral, dental, contact lens or body-care
XX  compositions) where they remove mannan-containing soils and prevent
XX  binding of some soils to cellulosics; and in fabric softeners. They
XX  can also be used in oil well drilling to fracture subterranean
XX  formations.
XX  Sequence 1470 BP; 441 A; 237 C; 355 G; 437 T; 0 other;
SQ

```

```

Query Match          100.0%; Score 900; DB 21; Length 1470;
Best Local Similarity 100.0%; Pred. No. 1.4e-261;
Matches 900; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  181 caatgcaaaaagatgacatccatagaagaacctatctcttgcgggaataat 240
DB  271 caatgcaaaaagatgacatccatagaagaacctatctcttgcgggaataat 330
QY  241 cattgtgtgtctgtcttgaagaattcatgatgtacacgtgttatgtatcttcgtc 300
DB  331 cattgtgtgtctgtcttgaagaattcatgatgtacacgtgttatgtatcttcgtc 390
QY  301 aatcgtctgttgtatatttgatgaagaatgaaagtgtcttaatttgaaagaagatacc 360
DB  391 aatcgtctgttgtatatttgatgaagaatgaaagtgtcttaatttgaaagaagatacc 450
QY  361 gtcattataatattgcgaatgaatggtttgttcgttggaagggaagccttggtctac 420
DB  451 gtcattataatattgcgaatgaatggtttgttcgttggaagggaagccttggtctac 510
QY  421 ggtataaacaagcaatcccgatctgtgtaacgcggtctcaaacacatccttggtga 480
DB  511 ggtataaacaagcaatcccgatctgtgtaacgcggtctcaaacacatccttggtga 570
QY  481 gatgtcgggggtggaagaattccacaatcgatctatgatgatbgaagaagtttt 540
DB  571 gatgtcgggggtggaagaattccacaatcgatctatgatgatbgaagaagtttt 630
QY  541 aatgtgcacctcaacgaataataatgtttcgattcatgatgatgaatgcagggtgt 600
DB  631 aatgtgcacctcaacgaataataatgtttcgattcatgatgatgaatgcagggtgt 690
QY  601 aatgatcgaagtctgactaataatttgaccgaagttcttaatacgaactcgcattagtc 660
DB  691 aatgatcgaagtctgactaataatttgaccgaagttcttaatacgaactcgcattagtc 750
QY  661 attgtgaatttggaacacgctcatacaatgtgtacgtgcataagcaacgattatgagc 720
DB  751 attgtgaatttggaacacgctcatacaatgtgtacgtgcataagcaacgattatgagc 810
QY  721 tattctgcaaaaagagaggttggtgtgttcgctggtcatgaaagggaagggccgga 780
DB  811 tattctgcaaaaagagaggttggtgtgttcgctggtcatgaaagggaagggccgga 870
QY  781 tgggaattttaagaccttcgaatgattggcttggaataaaccttacagcttgggaat 840
DB  871 tgggaattttaagaccttcgaatgattggcttgggaataaaccttacagcttgggaat 930
QY  841 acaataggaatgttcataatggtttaagaagaacctcgagataagcagcgttttaca 900
DB  931 acaataggaatgttcataatggtttaagaagaacctcgagataagcagcgttttaca 990

RESULT 2
ID 245336
ID 245336 standard; DNA; 1438 BP.
XX 245336;
XX 27-MAR-2000 (first entry)
DE DNA encoding a mannanase-linker-cellulose binding domain fusion.
XX
XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;
XX 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
XX galactoglucomannan; cellulosic fibres; synthetic fibre; yarn; fabric;
XX printing paste; plant material degradation; recycled waste paper;
XX paper making pulp; guar; locust bean gum; thickener; viscosity;
XX mannan-containing food; coffee extract; cleaning composition;
XX machine washing; hard-surface cleaner; dishwashing; oral; dental;
XX contact lens; body-care composition; fabric softener; oil well drilling;
XX subterranean formation fracture; cellulose binding domain; ss.
XX
XX Synthetic.
XX Bacillus sp.
XX Clostridium thermocellum.
OS

```


FT /product- "Mannanase"
 FT /EC_number- "3.2.1.78"
 FT /function- "Hydrolysis of 1,4-beta-D-mannosidic linkages"
 FT sig_peptide 1..93
 FT mat_peptide 94..1404
 FT /tag- b
 FT /tag- C
 FT /label- Mature_Mannanase
 XX W09964552-A1.
 XX
 XX 16-DEC-1999.
 PD
 PF 10-JUN-1998; 98WO-US12026.
 XX
 PR 10-JUN-1998; 98WO-US12026.
 XX
 PA (PROC) PROCTER & GAMBLE CO.
 XX
 PI Belliol JP, Thoen CAJK;
 XX
 DR WPI: 2000-116536/10.
 DR P-PSDB: Y44496.
 XX
 PI Detergent composition for removing greasy stains such as cosmetics,
 PT food stains and body soils
 XX
 PS Disclosure: Page 102-103; 113pp; English.
 XX
 CC The present sequence is the B. agaradherens Clone MB594 DNA encoding
 CC alkaline mannanase enzyme. It is derived from B. agaradherens strain
 CC NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
 CC from 7.5-10.5. It can be used in a detergent composition along with a
 CC mid-branched anionic surfactant. The detergent composition may be used
 CC for washing purposes, soaking/pre-treatment of stained fabric, hard
 CC surface cleaning and for removal of cosmetic and/or food stains. This
 CC composition provides excellent cleaning effect at low temperature.
 CC
 XX Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other:

Query Match 52.0%; Score 46812; DB 21; Length 1407;
 Best Local Similarity 70.1%; Pred. No. 2e-131;
 Matches 629; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

QY 4 gcaaatccggaattatgtaagcggtaccactctacagatgcaatggaaccattt 63
 DB 94 gcaagttacagcgctttatgtatggaatacgttatatgacgcaaatgagcaccatt 153
 QY 64 gtaatgaggggattacacatgagcgacgcatgtataaagacacgactctgcaatt 123
 DB 154 gtcagagaggtattacacatgacatgcttgttataaagacacgcttcaacagattt 213
 QY 124 gaaggagattgcaaatccggtgtaatacagtcgcggaattgtttatctgagggagaca 183
 DB 214 cctgcattgcagaggaagcaacgcaacgattcgattgtttatcagatggcggtcaa 273
 QY 184 tggacaagaatgacatcatcagtaagaacctatctctttagcggaataatctt 243
 DB 274 tgggaagaagaacgacatctgcaacacatctgtaagtcattgagcttcgagacaataaa 333
 QY 244 ttggttcgttcattgaattcatgattacggttattaccccttcgttcgcatc 303
 DB 334 atggggtcgttcgttgaaatgcatgacgaggtcgcgattcgcgcagtgatttaatt 393
 QY 304 cgtctgttattatgtaattgaatgagaagtgctttaaattggaagaagataccgtc 363
 DB 394 cgaagcgttgattatttgtagtaagaatgagatgccttaccgtaagaagataccgtt 453
 QY 364 attataatatgcaatgaaatggtttgttcgttggaagggagagcttggtcgaaggg 423
 DB 454 attataacattgcaaaagcagtgtagtggaagttggagtgctcagcttggtgcagatggc 513

QY 424 tataacaagcaatcccgagattcgtaagccggttcaaacctacttattgtaagat 483
 DB 514 tatattgtatgcattccgaagcttcgcgtatccggttcaaacacacacttattgtaagat 573
 QY 484 gctgggggtggggaattcccaacatgcatgattatgtaaggaagattttat 543
 DB 574 gcaagcagatgggggaatcccgcaatctatctatctatgtaaggaagatggtttat 633
 QY 544 gctgacctcaacgaataacataatgttttcgattcatatgattgaatgacgtgtaagat 603
 DB 634 gcaatccgttaaaaaatacagatgtctccatccatcatgattgattgctggtgtaagat 693
 QY 604 gcatcgcaagttcgttactaatattgaccgaagttcttcaatcaagccttcgattgtaagat 663
 DB 694 gtaaacactgttagatcaaatattgtatagatcagatcagaccttcgcttcgttata 753
 QY 664 ggtgaatttggacacgcgttcatacaaatgtgtaagcgtgtaagcaagcattatggtcat 723
 DB 754 ggtgaattcgttcataagacatactatgattgtgtgtgtaagataacataccttattat 813
 QY 724 tctgaacaagaaggaggttgggtgtggtgctggtatgtaaggaaggaagcgcagaaatg 783
 DB 814 tctgaagaagaactgacagaggtggtcgttcgttcttgtaaggaagcaacagatcagaaatg 873
 QY 784 ggtatttagaccccttcgaatgattggtgctggaataacacttaccgcttggtggaataga 843
 DB 874 gactatttagaccttcgaagacgtggtcgttcgaacttcaacattgattgggggaataga 933
 QY 844 atagtgaaatgctccatattgattgaagaagaactcgaagattgaagcagcttttaca 900
 DB 934 attgcccaggggcagatggtgttaagagaacactcacaacatccacgctatttaca 990

RESULT 4
 245338
 245338 standard; DNA; 1407 BP.

AC 245338;
 XX

DT 27-MAR-2000 (first entry)
 XX

DE DNA encoding a Bacillus mannanase enzyme.
 XX

KM Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KM endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KM mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KM plant material degradation; recycled waste paper; paper making pulp;
 KM guar; locust bean gum; mannan-containing food; coffee extract;
 KM cleaning composition; machine washing; hard-surface cleaner;
 KM dishwashing; oral; dental; contact lens; body-care composition;
 KM fabric softener; oil well drilling; subterranean formation fracture; ss.

OS Synthetic.
 OS Bacillus agaradherens.
 XX

FE Key location/Qualifiers
 FT CDS 1..1407
 FT /tag- a
 FT /product- "mannanase"

PN W09964619-A2.
 PD 16-DEC-1999.
 XX

PF 10-JUN-1999;
 XX

PF 99WO-DK00314.
 XX

PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.


```

PR 05-MAR-1999; 99DK-0000306-
PR 05-MAR-1999; 99DK-0000307-
PR 05-MAR-1999; 99DK-0000308-
PR 05-MAR-1999; 99DK-0000309-
PR 05-MAR-1999; 99DK-0000309-
PR 10-MAR-1999; 99DS-0123543-
PR 10-MAR-1999; 99DS-0123623-
PR 10-MAR-1999; 99DS-0123641-
PR 11-MAR-1999; 99DS-0123642-
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PI WPI: 2000-105891/09.
XX P-PSDB: Y54125.
XX
XX New mannases for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions -
XX
XX Example 5: Page 214-215; 242pp: English.
XX
XX The present sequence encodes a Bacillus mannanase (also known as mannan
XX endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase) .
XX The present mannanase is a beta-mannanase variant of the mannanase of
XX Y54124, in which the C-terminus of the protein was changed due to
XX design of a lower PCR primer used for amplification. The mannanase
XX hydrolyses galactomannans. Specifically, mannases hydrolyse
XX 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
XX glucamannans, and galactoglucomannans. The mannanase protein, or
XX preparations containing it, are used to improve properties of cellulosic
XX or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannases
XX are also used to process coffee extracts (to inhibit gel formation);
XX in cleaning compositions (for machine washing of fabrics, as
XX hard-surface cleaners, for hand or machine dishwashing, also in oral,
XX dental, contact lens or body-care compositions) where they remove
XX mannan-containing soils and prevent binding of some soils to
XX celluloses; and in fabric softeners. They can also be used in oil
XX well drilling to fracture subterranean formations.
XX
XX Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other:
SQ
Query Match 52.0%; Score 468.2; DB 21; Length 1407;
Best Local Similarity 70.1%; Pred. No. 2e-131.
Matches 629; Conservative 0; Mismatches 266; Indels 0; Gaps 0.
QY 4 gcaaatccggaatttattgtaagcgtgtacacatctatagacatggaatggaaccatt 63
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 gcaagttaacagcctttattgtttgtagtgcataagttatattatgacgcaaatggcagccatt 153
QY 64 gtaatgagaggatttaacacatgagcagcagatggtataaagaccaggacactctgcaatt 123
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 154 gtcatgagaggatttaacacatgagcagatggtttgataaagacacgccttcaacagctatt 213
QY 124 gaagagattgcaaaataacgcgtgtctaatacgcgcggatgtgtttcttctatggggagaca 183
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 214 cctgtccattgcaggaagcgcgcacacagatctgcatgttttctcagatggcggtcaa 273
QY 184 tgaacaaaagaatgacatccatcagaataaagaaacctatctctttagcggaaagatatcat 243
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 274 tgggaaaaagacgaattgacacacatttcgtgagtcatttgagcttgcggagacaaataaa 333
QY 244 ttgtgtctgttcttgaagttcaatgagtcacgcgtttatgattccattcgttcgcataat 303
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 334 atggtggtctgttcgttgaagttcatatgacgcagggctgcgagattcgcgcagtatttaaat 393
QY 304 cgtgtgtgtgatttattgattgaaatggaagtgttttaattggaaggaagatacgcgc 363
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 394 cgaagccttgattcatctgacataagaatgaaagtgcgtcttaccgtgttaaaagaagatacgcgt 453

```

QY	364	attatcaattacgtcgagatgaagatggtttgttcgttggaaggggaagcttggtgcgaagg	423
DQ	454	attcttaacctctcaaacgaagtggtctttggttggtgaagcttcgcttggtgcgaatgc <td>513</td>	513
QY	424	tataaacaagaacatccgcgcatctgcgtaacgccgcgtcttaaacacatcccttgatgtagat <td>483</td>	483
DQ	514	tatatgtatgtcattcccgaaagcttcgcggttcgcggttttaacacacacaccttaagtgtgat <td>573</td>	573
QY	484	gctcggggtgtgggacaattccacaatcgcattcattatgattatggaagagaagttttaat <td>543</td>	543
DQ	574	gcacgaggaatggygggcataatccgcacatcattcattcattacggaacaaagattggttaat <td>633</td>	633
QY	544	gctacccctcaacgaatacaatcgttttcgattcattatgtatgaaatgcaagtgtgtaat <td>603</td>	603
DQ	634	gcagatccgcttaaaaaatacagatgcttcctccatcccatatgtatgatatgctggtgtgat <td>693</td>	693
QY	604	gcacgcgaagtctgcgtactaataatgtgacgcgaatcttctaatacgaacccctgcattagtcatt <td>663</td>	663
DQ	694	gctaacacgtttgatgataaataatgtttagatgcatagatagaacagaccttgctctcgtaata <td>753</td>	753
QY	664	ggtcaatttggacaacggtcatatacaaatgtgtgacgttcgataggaagcaacgattatgacctat <td>723</td>	723
DQ	754	ggttaattcgcgtatagacatactatggttgatggtatgtatgaaagatacaatcccttgat <td>813</td>	813
QY	724	tcttaacaagaagagagttggtgtgtgtgcgtgcatctggaagaagggaagcccgacgaatgg <td>783</td>	783
DQ	814	tctgaagaagaacctgycacagaggtgctgcgcgtctggtcttgtaagaagcaacagatccgaatgg <td>873</td>	873
QY	784	gagattattagaccttcgaatgatagtgtggtctggaataaataccttacacgcttggtgggaataca <td>843</td>	843
DQ	874	gacctatttagaccttcogaagaagctggtgtgtgtaacattacattacgtatggtggggaataga <td>933</td>	933
QY	844	atagtgaatgtccatagtggtttaagaagaacctcgaggttaagaccgttttttca <td>900</td>	900
DQ	934	atgtgcacggyggccgcatgcttaccagaagaacctccaaacccatccaccgtaatttaca <td>990</td>	990
RESULT	5		
ID	229845		
AC	229845	standard; DNA; 1482 BP.	
XX	229845		
DT	27-MAR-2000	(first entry)	
DE	Bacillus agaradherens NCIMB 40482, Mannanase encoding DNA.		
XX			
KW	Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;		
KM	detergent composition; mid-branched anionic surfactant; washing;		
XX	cosmetic stain; food stain; ss.		
OS	Bacillus agaradherens.		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	1..1482	
FT		/*tag= a	
FT		/product= "Mannanase"	
FT		/EC_number= "3.2.1.78"	
FT		/function= "Hydrolysis of 1,4-beta-D-mannosidic linkages"	
FT	sig_peptide	1..96	
FT		/*tag= b	
FT	mat_peptide	97..1029	
FT		/*tag= C	
FT		/label= Mature_Mannanase	
XX			
XX	W09964552-A1.		
XX			
PD	16-DEC-1999.		
XX			
PF	10-JUN-1998;	98WO-US12026.	
XX	10-JUN-1998;	98WO-US12026.	

XX MPI: 2000-105891/09.
DR P-PSDB: Y54129.

PT New mannases for treatment of textiles, plant material and coffee
extract, and in cleaning compositions

XX Example 12, Page 223, 242pp: English.

XX The present sequence encodes a mannanase enzyme (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to cellulose; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.

XX Sequence 1107 BP; 363 A; 170 C; 255 G; 319 T; 0 other;

Query Match 45.4%; Score 408.4; DB 21; Length 1107;
Best Local Similarity 67.0%; Pred. No. 1.9e-113;
Matches 580; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

OY 13 ggaattatgtaagcgggtaccctctatcgatgccaatggaaccattgtaagaga 72
DB 217 ggttttatatagatggaacacccctatctatgtaacatggaacgattgtgctg 276
OY 73 gggatgaacatgagcagcatgataagaccgagcagcaccatgcaattgaaggatt 132
DB 277 ggaatgaacacggatcatcatcgtatcaagacctacatgaacacggcgatggaagcaatt 336
OY 133 gcaaataccgggtctaatatcaggtccggatgtgtatctcgtatggggagacaagaca 192
DB 337 gctgatactcgagacaaaccctcgtgtgagttctcttcgtagagacaaagggagca 396
OY 193 gatgacatccatcagtaagaaaccttatctcttagcggagaatacattgtgtcgt 252
DB 397 gatgatgtgacgaagtagcaaaaatlatatctttagcagaaaacatctttagtct 456
OY 253 gttcttgaagttcagatgctacggttatgattccatgctgctgctcaatgtgctt 312
DB 457 gctcttgagtcacatgacatgctcggaaagatgatatgtgaaccattactaaagct 516
OY 313 gattatggattgaaatggaagtgcttaatttgaaagaaagataccgtcatat 372
DB 517 gattactgattgagatcaagaagatcttaactcgaaagagacaagtaattattac 576
OY 373 attggaatgaaatgtgtgtgttcggtggaagggatgctgtgctggaagggatataca 432
DB 577 attctaatgaaatgtgtgtgttcggtggaagggatgctgtgctggaagggatataca 636
OY 433 gcaatcccgatgtgtaagcggcgctcaaaccaaccttatgtatgattcgggg 492
DB 637 gcaatcccttctcgaagagagggggtcttaaacataccttaatgtgacgacgtgg 696
OY 493 tggggacaattccacaatcgtatcgtatctatgaaagagaatttttaattcgtacct 552
DB 697 tggggacaattccctcagatctatcatgaaaaagattgaagtttttaaccagaccca 756
OY 553 caacgaataacatgttttgcattatgtatgataatgacaggtgtgtaatgacgaca 612

DB 757 ttaagaataacatggtttccatcatcatatgtatgaatggcaggggtatcctcaaca 816
OY 613 gtctgactaatattgaccgagttcttaatacaagacctgcattagtcattgtaatt 672
DB 817 gtaaaagacaatattgacggtgtctcttgaagaagatttagttagtaattgtagtgc 876
OY 673 ggaacgcttacaatagtgacgtcgtatgaagaacagatattagatctcgaacaa 732
DB 877 gttcatctccactacggaagagatcgtgtgtatatacaacttaagtcattcgtaga 936
OY 733 agagagttggt 792
DB 937 tatgtatgagttgt 996
OY 793 gactcttgaatgattggt 852
DB 997 gactatgaaacagatcttcccgagacacactaactgaatggtggaagaattgtac 1056
OY 853 ggtcatatggtttaagaagaacttc 878
DB 1057 gttccgaatgttttaagaagaacttc 1082

RESULT 8
ID 245340 standard; DNA; 995 BP.
XX 245340;
AC 245340;
XX 27-MAR-2000 (first entry)
DE DNA encoding a *Bacillus mannanase* enzyme.
XX
KW Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture; ss.
XX
OS *Bacillus* sp.
XX
EH Key Location/Qualifiers
FT CDS 1..993
FT /tag- a
FT /product= "mannanase"
FT /note= "not termination codon given"
FT sig_peptide 1..96
FT /tag- b
XX
XX W09964619-A2.
XX
PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99MO-DK00314.
XX
XX 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 03-MAR-1999; 99DK-00001725.
PR 03-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.

XX (NOVO) NOVO-NORDISK AS.
 PA Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 PI MPI: 2000-105891/09.
 XX P-PSDB: Y54127.
 DR New mannases for treatment of textiles, plant material and coffee
 XX extract, and in cleaning compositions -
 PT Example 10; Page 219-220; 242pp; English.
 PS
 XX The present sequence encodes a *Bacillus mannanase* (also known as
 CC Mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 CC
 SQ Sequence 995 BP; 286 A; 162 C; 256 G; 291 T; 0 other;

Query Match 45.0%; Score 405; DB 21; Length 995;
 Best Local Similarity 66.1%; Pred. No. 1,9e-112;
 Matches 585; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

QY 13 ggaattatgtaagcgggtaccactctatagatgcaaggaacccattgtatgaga 72
 DB 106 ggggtcattatgtaagatgatacattatgatacgaatcccttgtatgaga 165
 QY 73 gggattaacatggcgatggtataagacagcgaactctgtaattgaaagatt 132
 DB 166 gggattacatgagcgtctgtttaacaagaactaataacatcagagggagatt 225
 QY 133 gcaataccggtcctaatacagtcggaattgtatctgatagggagacaacaaa 192
 DB 226 agtcaaacaggggacaatacagtcggtctgtcttaatgaggaagatgcaaaaa 285
 QY 193 gatgacatcacaacagtaagaacattatccttagcggagaataatcattgtgtct 252
 DB 286 gatgattgaaagatgagcttcgttattcttggcagagacgacataaattgtcc 345
 QY 253 gtctctgaattatgctgtacccggtatgattccattcttcgtcccaatcggtct 312
 DB 346 gtctttagaagatgactgtactgttagaataattcttcgactgcgaagctgctg 405
 QY 313 gattattggaatgaagagagagtccttaattgaaaggagatacgcattatnaat 372
 DB 406 gactattggaatgagagagagatgttttcagagggaagaagacataagtgatcataat 465
 QY 373 atcgcaaatgaatggttgcgttcgtaggaaggagagtcgttgagcgggtataaaca 432
 DB 466 atcgcaaatgaatggttgcgttcgtaggaaggagcgcagtgagaggtatcagaat 525
 QY 433 gcaatcccgagatgctgtaacgcgcgttaaacataccttgatgagatgagcgggg 492
 DB 526 ggcgtacagtcagctcgaaatgcaggcttgcacatacatttaagtgacgcgtcggt 585
 QY 493 tggggacaattccacaatcgatcagattatggaagagagattttaatgctgaccc 552

DB 586 tatggccagtaaccctcaatcagtgatgtatgtatgtaagaagtattaaatgctgacca 645
 QY 553 caacgaatacacaatgtttcttcgattcataatgtatgataatcagtgatgcaac 612
 DB 646 cagaagaacacaatgtttcttcgattcataatgtatgataatcagtgatgcaac 705
 QY 613 gtctgtaactaatgtgacagtgatccttaacagaccccgatgattgtgtaatt 672
 DB 706 gtaagacgaacacattgctgctgactgaaccgaacattgactctgtcattgtgtaatt 765
 QY 673 ggaacacgcatacaaatgtgacgctgcagtaagaagcaacagatataagcttgaacaa 732
 DB 766 gggcattgcatatgacggttgatgtatgagagacccatttaagctattccagcaa 825
 QY 733 agagagatggtgtgtgtgctgtgcatgaaagggaggaacagtgagatatta 792
 DB 826 agaatgtgagatgtgtgtgtgagctgcatgacatagtgaaagggcagataact 885
 QY 793 gactctggaatgattggtgtggaataacattacagctgggggaataatagatgaat 852
 DB 886 gattacgaatgactctgtgtgtaacgacatgacatgctggtgggtgacataagtaac 945
 QY 853 gtcacatgtttaagaagaacttcgagattagacacgctttt 897
 DB 946 gttccggaatgtgattgtctcaacactcaagaagacagtggttt 990

RESULT 9
 245345
 ID 245345 standard; DNA; 960 BP.
 XX
 AC 245345;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE DNA encoding a *Bacillus* sp. mannanase enzyme.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture; ss.
 XX
 OS *Bacillus* sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..960
 FT /*tag- a
 FT /product- "mannanase"
 FT /note- "not termination codon given"
 FT sig_peptide 1..84
 FT /*tag- b
 XX
 PN MO9964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PE 10-JUN-1999; 99MO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.

PR 05-MAR-1999: 99DK-0000308.
 PR 05-MAR-1999: 99DK-0000309.
 PR 09-MAR-1999: 99US-0123543.
 PR 10-MAR-1999: 99US-0123623.
 PR 10-MAR-1999: 99US-0123641.
 PR 11-MAR-1999: 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 DR P-PSDB: Y54133.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 XX
 PS Disclosure; Page 229; 242pp; English.
 XX

The present sequence encodes a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulose; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 CC
 XX Sequence 564 BP; 159 A; 111 C; 158 G; 136 T; 0 other;

Query Match 18.6%; Score 167.8; DB 21; Length 564;
 Best Local Similarity 59.2%; Pred. No. 4.8e-41;
 Matches 286; Conservative 0; Mismatches 197; Indels 0; Gaps 0;
 QY 3 tgcgaattccgaatttattgaagcggtaccactatacagatgcaatggaaacccatt 62
 DB 81 tgcacaaagcggtcttcacgtaaaaggtacagagttgttgacaaaatgsgatcctta 140
 QY 63 tgcgaagagcggtatccatcggtgcacgcatggtataaagaccaggaactactgcaat 122
 DB 141 cgttatcggtggtgcacacatgacattcttggtttaacaaagatttaagagagcaat 200
 QY 123 tgaagagattgcaaatccggtcttaatacagtcgcgattgtgttatcgtatggagaca 182
 DB 201 cccctgcacatgcagaaacagggcgacacacagtgagatcgtcttaacatgacagca 260
 QY 183 atggacaaaagatgacatccatcagtaagaacattatcttctttagcggagaatataca 242
 DB 261 atggagaaaagatgacatcccttcgctgcgtgcgtgtgcttgcacagaaacatatg 320
 QY 243 ttggtgtcgtcttgaagttatcgttgcacggtatgattcattccttgcgtcaa 302
 DB 321 gttgacaaacggtctggaagttccacgattctacaggaagtgataatcccatgattaga 380
 QY 303 tctgctgttattgattgattgaatgagaagtgcttattattgaaaggaatgatacgt 362
 DB 381 taagacagtcgattactgcgtgcaaatgctgattcttcaaggggacaaagccggt 440
 QY 363 catattaatattgcaaatgaaatggttggctcgtggaaaggaggtccttggtgacg 422
 DB 441 aatcattaatcattgcgaatgattgtatggggcggtgagagtgatgactttgggcaaaagc 500

QY 423 gataaacaagcaatcccggtattgctgaacggcgtcttaacacattcgtatgtaga 482
 DB 501 atacgcacaagcgtatcccggtctgagcagtgcttgcctgcacatacgttaataatga 560
 QY 483 tgc 485
 DB 561 tgc 563

RESULT 11
 245344
 ID 245344 standard; DNA; 397 BP.
 XX
 AC 245344;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE DNA encoding a partial Bacillus mannanase enzyme.
 XX

Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture; ss.
 XX
 OS Bacillus clausii.
 XX

Key Location/Qualifiers
 FT 1..396
 FT /tag- a
 FT /product- "mannanase"
 FT /note- "partial sequence"
 XX

W09964619-A2.

16-DEC-1999.

10-JUN-1999: 99NO-DK00314.

PR 10-JUN-1999: 99US-0111256.
 PR 20-OCT-1998: 98DK-0001340.
 PR 20-OCT-1998: 98DK-0001341.
 PR 28-OCT-1998: 98US-0105970.
 PR 28-OCT-1998: 98US-0106054.
 PR 23-DEC-1998: 98DK-0001725.
 PR 05-MAR-1999: 99DK-0000306.
 PR 05-MAR-1999: 99DK-0000307.
 PR 05-MAR-1999: 99DK-0000308.
 PR 05-MAR-1999: 99DK-0000309.
 PR 09-MAR-1999: 99US-0123543.
 PR 10-MAR-1999: 99US-0123623.
 PR 10-MAR-1999: 99US-0123641.
 PR 11-MAR-1999: 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX

PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 DR P-PSDB: Y54131.
 XX

PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 XX
 PS Disclosure; Page 226-227; 242pp; English.
 XX

The present sequence encodes a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannase hydrolyses galactomannans. Specifically, mannases


```

OY 616 cgtactaatattgaccgagttcttatacaagaacctcgacttagtcattggtgaatttga 675
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 178 ggcagcgaatttcaggccatcaagaaccttggcttgcgtcatgattggtgaattcga 237
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 676 caccgcttacaagaatg-----gtgacgttgatgaagcaacgattatg 717
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 238 tacaactaacaacgcaataacaacttggggaggttaacgcccaggaataacatg 297
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 718 agctatttcgaacaaagagaggttgggttggcgttgcattggaagaagcgccca 777
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 298 aatcagcgcgcaagcaaaaggaatcggtacatgcgctgtgctgagcttgcaatgacg 357
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 778 gaattggagatttagaccttcgaattcgaattggtgctggaataacctcagcttggga 837
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 358 gtaacctcttggttgatnagacacaaa---cgattggcaaacacttaccatcattcgtg 414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 838 aatacaatagtagtgcataatggttaagagaacttcgaatgaacacgctttt 897
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 415 aactagttgtaaatggaacacgcaatcgagctacgctcgtccagcaactgtattt 474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 898 a 898
    |
DB 475 a 475

RESULT 13
T29774
ID T29774 standard; DNA: 2040 BP.
XX
AC T29774;
XX
DT 19-NOV-1996 (first entry)
XX
DE Bacillus thuringiensis ssp. israelensis CryIVD protein DNA.
XX
KW CryIVD: toxic protein; crystal toxin; expression construct;
KW transformed cyanobacteria; phycocyanin beta; cpb; promoter;
KW insecticide; dipteran larvae; mosquito; blackfly; ss.
XX
OS Bacillus thuringiensis.
XX
FH Key Location/Qualifiers
FT CDS 41..1972
FT /tag= a
XX
PN US5518897-A.
XX
PD 21-MAY-1996.
XX
PE 04-MAY-1992; 92US-0877876.
XX
PR 28-JAN-1994; 94US-0188581.
PR 04-MAY-1992; 92US-0877876.
XX
PA (UTME-) UNIT MEMPHIS STATE.
XX
PI Murphy RC, Stevens SE;
XX
DR MPI: 1996-259063/26.
XX
DR P-PSDB: R97735.
XX
PT New DNA construct for expressing cryIV D protein in cyanobacteria -
PT under control of a phycocyanin beta promoter, useful for control of
PT dipteran larvae in water
XX
PS Example 1; Columns 9-14; 20pp; English.
XX
CC The present sequence encodes the B. thuringiensis ssp. israelensis
CC CryIVD toxic protein, which was used in the prepn. of a claimed DNA
CC construct for the expression of CryIVD in cyanobacteria, comprising
CC the present sequence under the control of phycocyanin beta (cpb)
CC promoter. Cyanobacteria (which may be adapted for growth in fresh

```

```

CC or brackish water) transformed with the construct can be used
CC as insecticides for controlling dipteran larvae, esp. those of
CC mosquitoes and blackflies, that live in water. The percentage of
CC Culex pipiens (mosquito) larvae surviving after 4 days with
CC cyanobacteria transformed with the claimed DNA construct as their
CC only food source was 51 %, compared to 94 % for those fed with
CC cyanobacteria transformed with an empty plasmid. In the
CC cyanobacteria, CryIVD is efficiently expressed under the control of
CC the strong cpb promoter, even though the CryIVD gene contains 19
CC AUA which are generally poorly translated (if at all) in
CC cyanobacteria.
XX
SQ Sequence 2040 BP; 716 A; 315 C; 373 G; 636 T; 0 other;

Query Match      3.9%; Score 35; DB 17; Length 2040;
Best Local Similarity 47.7%; Pred. No. 0.92;
Matches 135; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

OY 91 gcatggtataaagaccaggaactactgcgaattgaaggattgcaaatccggtgctaat 150
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1355 gcatataatcagcatcttacaacactaacataatagatagatgattcactact 1414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 151 acggtccggattgtgtatctatgtaggggagacaatgagcaaaagatgacatcatagta 210
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1415 acagaataatattatgtagggtttgacacagataataactcaaggactttattcctaaana 1474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 211 agaaacctatctctttagcgaagaataatcattggttcgtctcttgaattcatgat 270
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1475 tctcatatttaagtaagaacgaattgtagttatgtatccgtctcgaatttgcgtga 1534
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 271 gctacgg--gttatgattccattgcttcgtcgaatcgtgctgtgattgaattga 327
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1535 gtttcagatagatcatttttagaagatacgcagatcaagcaacgacgagattaa 1594
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 328 atgagagtgctttaatttgaaaggagataccgtcatatta 370
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1595 ttgcacgtacttccattagtagtaagctaaagtactcatca 1637
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
N90712
ID N90712 standard; DNA: 2100 BP.
XX
AC N90712;
XX
DT 09-JAN-1990 (first entry)
XX
DE cryD gene.
XX
KW cryD protein; Bacillus thuringiensis; biopesticide.
XX
DE Bacillus thuringiensis var. israelensis.
XX
OS W08907605-A.
XX
PN 24-AUG-1989.
XX
PD 17-FEB-1989; 89WO-US00663.
XX
PF 19-FEB-1988; 88US-0158176.
XX
PR (ECOG ) ECOGEN INC.
XX
PA Donovan WP;
XX
DR MPI: 1989-263682/36.
XX
DR P-PSDB: P91462.
XX
PT Bacillus thuringiensis var israelensis cry D toxin gene and proteins
PT - used for producing insecticide compsns.active against Dipteran species.
XX
PS Claim 1; fig 2; 58pp; English.

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:56:10 ; Search time 75.74 Seconds
(without alignments)
1797.160 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_990
Perfect score: 900
Sequence: 1 aatgcaaatccgatttta.....gattagcaccgttttaca 900

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCrus.COMB.seq:*
7: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.4	4.2	7218	1 US-08-232-463-14	Sequence 14, Appl
2	33.4	3.7	1365	4 US-08-870-827-4	Sequence 4, Appl
3	33.4	3.7	2408	4 US-08-870-827-5	Sequence 5, Appl
4	33	3.7	1800	5 US-09-039-773A-1	Sequence 1, Appl
5	32	3.6	4359	3 US-08-566-398-23	Sequence 23, Appl
6	32	3.6	4365	6 PCT-US91-08525-21	Sequence 21, Appl
7	32	3.6	4365	6 PCT-US91-08525-25	Sequence 25, Appl
8	32	3.6	4365	6 PCT-US93-04384-1	Sequence 1, Appl
9	32	3.6	4365	6 PCT-US93-04384-7	Sequence 7, Appl
10	32	3.6	4365	6 PCT-US93-04384-15	Sequence 15, Appl
11	32	3.6	6090	3 US-08-566-398-35	Sequence 35, Appl
12	32	3.6	6144	3 US-08-566-398-32	Sequence 32, Appl
13	31.8	3.5	7366	7 5169760-3	Patent No. 5169760
14	31.8	3.5	7852	5 US-08-836-022A-2	Sequence 2, Appl
15	31.8	3.5	7897	5 US-08-836-022A-1	Sequence 1, Appl
16	31.8	3.5	8509	2 US-08-462-014-1	Sequence 1, Appl
17	31.8	3.5	10398	3 US-08-331-384-1	Sequence 1, Appl
18	31.8	3.5	10398	3 US-08-708-188-1	Sequence 1, Appl
19	31.8	3.5	10398	3 US-08-836-087-1	Sequence 1, Appl
20	31.8	3.5	19182	4 US-08-850-880-11	Sequence 11, Appl
21	31.8	3.5	19182	4 US-08-946-916-11	Sequence 11, Appl
22	31	3.4	1453	4 US-08-252-995D-1	Sequence 1, Appl
23	31	3.4	1453	4 US-08-834-108-1	Sequence 1, Appl
24	31	3.4	1600	4 US-08-252-995D-5	Sequence 5, Appl
25	31	3.4	1600	4 US-08-834-108-5	Sequence 5, Appl
26	31	3.4	2882	1 US-08-424-567-1	Sequence 1, Appl

27	31	3.4	2882	3 US-08-711-928-1	Sequence 1, Appl
28	31	3.4	3447	1 US-08-252-995D-3	Sequence 3, Appl
29	31	3.4	3447	4 US-08-834-108-3	Sequence 3, Appl
30	30.6	3.4	18994	2 US-08-459-586-4	Sequence 4, Appl
31	30.6	3.4	18994	2 US-08-282-696-4	Sequence 4, Appl
32	30.2	3.4	5718	5 US-08-714-918-48	Sequence 48, Appl
33	30	3.3	2675	1 US-07-749-001-4	Sequence 4, Appl
34	30	3.3	2675	1 US-08-154-158-4	Sequence 4, Appl
35	30	3.3	2675	1 US-08-463-335-4	Sequence 4, Appl
36	30	3.3	2675	3 US-08-464-023A-4	Sequence 4, Appl
37	30	3.3	5319	2 US-08-169-927-1	Sequence 1, Appl
38	29.8	3.3	1029	5 US-08-978-456-1	Sequence 1, Appl
39	29.8	3.3	3336	3 US-08-978-456-7	Sequence 7, Appl
40	29.8	3.3	3336	3 US-08-978-456-7	Sequence 7, Appl
41	29.8	3.3	4467	2 US-08-565-907A-1	Sequence 1, Appl
42	29.8	3.3	4467	4 US-08-910-551B-1	Sequence 1, Appl
43	29.8	3.3	4467	4 US-08-909-4425A-1	Sequence 1, Appl
44	29.8	3.3	51952	5 US-08-947-823-1	Sequence 1, Appl
45	29.6	3.3	1716	5 US-09-029-267-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTgpt-Fls
US-08-232-463-14


```

RESULT 6
PCT-US91-08525-21
; Sequence 21, Application PC/TUS9108525
GENERAL INFORMATION:
APPLICANT: SmithKline Beecham, Corporation
TITLE OF INVENTION: Recombinant feline Coronavirus S
TITLE OF INVENTION: Proteins
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08525
FILING DATE: 19911114
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: King, William T.
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: SBC 14532B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US91-08525-21

Query Match 3.6%; Score 32; DB 6; Length 4365;
Best Local Similarity 65.3%; Pred. No. 2.6;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 520 gattatggagaagaatttttaatcgtgacccctcaacgaatacaatggttcgattcat 579
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2593 GATTGTGCAAGATAGCTTGTGTAATGTAACCTAGATGTAAACAATTTTAACACAAATAT 2652
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 580 atgtatgcaatat 591
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2653 GTGTCTGCATGT 2664

RESULT 7
PCT-US91-08525-25
; Sequence 25, Application PC/TUS9108525
GENERAL INFORMATION:
APPLICANT: SmithKline Beecham, Corporation
TITLE OF INVENTION: Recombinant feline Coronavirus S
TITLE OF INVENTION: Proteins
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08525
FILING DATE: 19911114
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: King, William T.
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: SBC 14532B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US91-08525-21

```

```

1 STREET: 709 Swedeland Road
2 CITY: King of Prussia
3 STATE: PA
4 COUNTRY: USA
5 ZIP: 19406-2799
6
7 COMPUTER READABLE FORM:
8
9 MEDIUM TYPE: Floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: Patent Release #1.0, Version #1.25
13
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: PCT/US91/08525
16 FILING DATE: 19911114
17 CLASSIFICATION: 435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US 07/698,927
20 FILING DATE: 13-MAY-1991
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 07/613,066
23 FILING DATE: 14-NOV-1990
24 ATTORNEY/AGENT INFORMATION:
25 NAME: King, William T.
26 REGISTRATION NUMBER: 30,954
27 REFERENCE/DOCKET NUMBER: SPC 14532B
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (215) 270-5015
30 TELEFAX: (215) 270-5090
31 INFORMATION FOR SEQ ID NO: 25:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 4365 base pairs
34 TYPE: NUCLEIC ACID
35 STRANDEDNESS: double
36 TOPOLOGY: unknown
37 MOLECULE TYPE: CDNA
38 FEATURE:
39 NAME/KEY: CDS
40 LOCATION: 1..4362
41 PCT-US91-08525-25
42
43
44 Query Match 3.6%; Score 32; DB 6; Length 4365;
45 Best Local Similarity 65.3%; Pred. No. 2.6;
46 Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps
47
48 QY 520 gattatggaagaaaggttttaatgtcaccctcaacgaataacatgtttcgatcatt 579
49 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 Db 2553 gatttctgcaagctagctgtgtaattgtaaccctagatgtaacaaattgtttacacaaat 2652
51 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 QY 580 atgtatgaat 591
53 ||||| |||||
54 Db 2653 gttcttcgactgt 2664
55
56
57 RESULT 8
58 PCT-US93-04384-1
59 Sequence 1, Application PC/TUS9304384
60 GENERAL INFORMATION:
61 APPLICANT: Miller, Timothy J.
62 APPLICANT: Klepfer, Sharon
63 APPLICANT: Reed, Albert Paul
64 APPLICANT: Jones, Elaine V.
65 TITLE OF INVENTION: Compositions and Methods for Vaccination
66 TITLE OF INVENTION: Against Coronaviruses
67 NUMBER OF SEQUENCES: 48
68 CORRESPONDENCE ADDRESSES:
69 ADDRESSEE: SmlnKline Beecham Corporation - Corporate
70 ADDRESSEE: Patents
71 STREET: 709 Swedeland Road
72 CITY: King of Prussia
73 STATE: PA
74 COUNTRY: USA
75 ZIP: 19406-2799
76
77 COMPUTER READABLE FORM:
78

```


Query Match	3.68;	Score 32;	DB 6;	Length 4365;
Best Local Similarity	65.38;	Pred. No. 2.6;		
Matches 47;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;

RESULT 10
PCT-US93-04384-15

1 GENERAL INFORMATION:
 2 APPLICANT: Miller, Timothy J.
 3 APPLICANT: Klepfer, Sharon
 4 APPLICANT: Reed, Albert Paul
 5 APPLICANT: Jones, Elaine V.
 6 TITLE OF INVENTION: Compositions and Methods for Vaccination
 7 TITLE OF INVENTION: Against Coronaviruses
 8 NUMBER OF SEQUENCES: 48
 9 CORRESPONDENCE ADDRESS:
 10 ADDRESSEE: SmithKline Beecham Corporation - Corporate
 11 ADDRESSEE: Patents
 12 STREET: 709 Swedeland Road
 13 CITY: King of Prussia
 14 STATE: PA
 15 COUNTRY: USA
 16 ZIP: 19406-2799
 17 COMPUTER READABLE FORM:
 18 MEDIUM TYPE: Floppy disk
 19 COMPUTER: IBM PC compatible
 20 OPERATING SYSTEM: PC-DOS/MS-DOS
 21 SOFTWARE: PatentIn Release #1.0, Version #1.25
 22 CURRENT APPLICATION DATA:
 23 APPLICATION NUMBER: PCT/US93/04384

FILING DATE: 19930507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,171
FILING DATE: 08-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SRC H85009-1
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US93-04384-15

Query Match
Best Local Similarity 65.3%; Score 32; DB 6; Length 4365;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 520 gattatgaaagaagtttttaatgctgacccccaacgaataatgtttcgattcat 579
|||||
Db 2593 GATTGTGCAAGATACGTTGTAAAGTAACCTAGATGTACAAATGTTAACAATAT 2652

QY 580 atgatgaatat 591
|||||
Db 2653 GTGTCTGCATGT 2664

RESULT 11
US-08-566-398-35
Sequence 35, Application US/08566398
Patent No. 5858373
GENERAL INFORMATION:
APPLICANT: Paolletti, Enzo
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIOUS
PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
TITLE OF INVENTION: MAKING AND USING THEM
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/566,398
FILING DATE: 01-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2880
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 6090 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-566-398-35

Query Match
Best Local Similarity 65.3%; Score 32; DB 3; Length 6090;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 520 gattatgaaagaagtttttaatgctgacccccaacgaataatgtttcgattcat 579
|||||
Db 3123 GATTGTGCAAGATACGTTGTAAAGTAACCTAGATGTACAAATGTTAACAATAT 3182

QY 580 atgatgaatat 591
|||||
Db 3183 GTGTCTGCATGT 3194

RESULT 12
US-08-566-398-32
Sequence 32, Application US/08566398
Patent No. 5858373
GENERAL INFORMATION:
APPLICANT: Paolletti, Enzo
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIOUS
PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
TITLE OF INVENTION: MAKING AND USING THEM
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/566,398
FILING DATE: 01-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2880
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 6144 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-566-398-32

Query Match
3.6%; Score 32; DB 3; Length 6144;

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/331,381
FILING DATE: 28-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNVPR,008PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7897 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-836-022A-1

Query Match 3.5%; Score 31.8; DB 5; Length 7897;
Best Local Similarity 46.9%; Pred. No. 4.1;
Matches 99; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 61 ttgtaatgagaggaattacacatggtgacgcatggtataaagaccaggcaactactgca 120
DB 870 TTGTCTAGAAAGACAGCAACAGCGGTGGAATATAGCGAGAAAGAAATAGCGGC 929
QY 121 attgaaggatcgcaaatccggtgcttaataaggtccgagattggttactgagtgagga 180
DB 930 AAAAAAATATACCCGTATCTTTGCTGATATGTTGATGTCATGTAGCCAAATCGGGA 989
QY 181 caatgacaaagaatgacatcacatagtaagaacattatctcttagcggagaagataat 240
DB 990 AAACGGAGTAGGCTCCCATGATATAAAGTAAGAAAGAAATTAACCGACATCCA 1049
QY 241 catttggtgctgctctgaagttcatgaty 271
DB 1050 AAGTTTGTGTTTTTTAAATAGTACATAATG 1080

Search completed: December 20, 2000, 09:56:44
Job time: 30791 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 01:58:22 ; Search time 1967.85 Seconds

(without alignments)
2827.718 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_990

Perfect score: 900

Sequence: 1 aatgcaaatccgatttta.....gattaagcacgcgtttta 900

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_esthum21:*
64: em_esthum22:*
65: em_esthum23:*
66: em_esthum24:*
67: em_esthum25:*
68: em_esthum26:*
69: em_esthum27:*
70: em_esthum28:*
71: em_esthum29:*
72: em_esthum30:*
73: em_esthum31:*
74: em_esthum32:*
75: em_esthum33:*
76: em_esthum34:*
77: em_esthum35:*
78: em_esthum36:*
79: em_esthum37:*
80: em_esthum38:*
81: em_esthum39:*
82: em_esthum40:*
83: em_esthum41:*
84: em_esthum42:*
85: em_esthum43:*
86: em_esthum44:*
87: em_esthum45:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

117: gb_gss18:*
 118: gb_gss19:*
 119: em_gss13:*
 120: gb_gss20:*
 121: gb_gss21:*
 122: gb_gss22:*
 123: gb_gss23:*
 124: gb_gss24:*
 125: em_gss14:*
 126: em_gss15:*
 127: em_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45.6	5.1	1101	121	CNS0039G
2	43.4	4.8	1101	121	CNS001R2
3	40.4	4.5	1101	121	CNS0039U
4	39.4	4.4	552	96	AQ332958
5	39.2	4.4	520	100	AQ621653
6	39	4.3	750	115	AZ138670
7	38.8	4.3	459	21	AM333544
8	38.8	4.3	515	21	AM333365
9	38.8	4.3	545	21	AM334819
10	38	4.2	250	15	AU060486
11	38	4.2	1101	121	CNS0039U
12	37.6	4.2	524	90	AQ204415
13	37.6	4.2	674	124	CNS0543D
14	37.6	4.2	1101	121	CNS001R2
15	37.4	4.2	510	97	AQ413085
16	37.4	4.2	545	98	AQ534365
17	37.4	4.2	601	96	AQ375742
18	37.4	4.2	717	116	AZ196394
19	37.4	4.2	1101	121	CNS0106X
20	37.4	4.2	1101	121	CNS017XD
21	37	4.1	510	25	AM943977
22	37	4.1	938	121	CNS0067J
23	36.8	4.1	521	115	AZ129098
24	36.8	4.1	941	123	CNS031PY
25	36.6	4.1	1101	121	CNS0019R
26	36.2	4.0	715	117	AZ257353
27	36.2	4.0	890	121	CNS017FS
28	36.2	4.0	900	36	C93831
29	36.2	4.0	1100	121	CNS0063S
30	36	4.0	391	22	AM472600
31	36	4.0	533	9	A1166437
32	36	4.0	550	15	AU053215
33	36	4.0	853	122	CNS023KH
34	36	4.0	929	121	CNS016AV
35	36	4.0	936	121	CNS046JN
36	36	4.0	1101	121	CNS0010C
37	36	4.0	1101	121	CNS002C3
38	35.8	4.0	1033	121	CNS0143L
39	35.8	4.0	1101	121	CNS001R2
40	35.6	4.0	1101	121	CNS000D1
41	35.4	3.9	389	3	AA372519
42	35.4	3.9	449	90	AQ207422
43	35.4	3.9	536	25	AM932995
44	35.4	3.9	562	121	CNS016OR
45	35.4	3.9	602	25	AM933405

ALIGNMENTS

RESULT 1
 CNS0039G/c

LOCUS	CNS0039G	1101 bp	DNA	GSS	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL063921				
VERSION	AL063921.1	GI:4941778			
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutomo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .				
FEATURES	Location/Qualifiers				
source	1..1101				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
	/clone_lib="RPCI-98"				
	/clone="BACR08K10"				
	/note="end : TET3"				
BASE COUNT	201 a	64 c	131 g	202 t	503 others
ORIGIN					
Query Match	5.1%; Score 45.6; DB 121; Length 1101;				
Best Local Similarity	14.3%; Pred. No. 0.028;				
Matches	76; Conservative 244; Mismatches 211; Indels 1; Gaps 1;				
OY	60 attgtaatgagaggaatccatcggtgacgcgtgtatgaagacacgaactactgc 119				
DB	1080 WTKKMTWTKKDRADRRKAGDADRAMDGACTWTATWMMWMAATWDMWMAAT 1021				
OY	120 aattgaaggattgcaaatccatcggtgctaaatcggtgattggtatcgtatgaggg 179				
DB	1020 AAKTDTMTWTRWTRWTRWTRWTRWTRWTRWTRWTRWTRWTRWTRWTRWTR 961				
OY	180 acaatgacacaaagatgactccatcaatgaagaacactatctctttagcgaagataa 239				
DB	960 DKKGKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRR 901				
OY	240 tcaattggttcgttccttcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaat 299				
DB	900 DTGCTKDDDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDD 841				
OY	300 caatcgctgctgattgattgattgattgattgattgattgattgattgattgattg 359				
DB	840 RWDAMWKKWDAMWKKWDAMWKKWDAMWKKWDAMWKKWDAMWKKWDAMWKKWD 791				
OY	360 cgtcatatataatgacgaatgattgattgattgattgattgattgattgattgattg 419				
DB	780 TTTTWTTRTTRTTRTTRTTRTTRTTRTTRTTRTTRTTRTTRTTRTTRTTRTTR 721				
OY	420 cgggtataacaag-caatcccgattggttaacgcggtcctaaccataacttgatg 478				

```

Db 720 TWMDDADBDTARDDRKRKGGDAGAGKKTGKRKRDRRTWTDWMDAAWMTT 661
Oy 479 tagatgctgcggggtgggacattccacatcattcatgatataggagaagatt 538
Db 660 DTDTDMDKRRRRRRRRRTTAAADWMTWKAMDAKMDKTRADRMWMDADTWT 601
Oy 539 ttaatgctgcacccaagaataacatttcttgattcatatgtatgata 590
Db 600 DARKADRWAKARAWRARARARADRRMTTKGTTATWTTWAAARAWW 549

RESULT 2
CNS00LT2 1101 bp DNA GSS 14-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR48P19 of RPCT-98 library from Drosophila melanogaster (fruit
AL078714 genomic survey sequence.
AL078714.1 GI:5102004
GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Location/Qualifiers
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1lb="RPCT-98"
/clone="BACR48P19"
/note="end : TET3"

BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN
Query Match 4.8%; Score 43.4; DB 121; Length 1101;
Best Local Similarity 14.9%; Pred. No. 0.12;
Matches 48; Conservative 140; Mismatches 135; Indels 0; Gaps 0;

Oy 225 tttagcgaagaatacattgttgcgtcttgaaagtcattgctaccggtatga 284
Db 699 TTKKKKRAAADAADKMDAKKAKATKTKKKKRAAADAADKDKRKGKKGKKG 758
Oy 285 ttccattgctgcacatgcgtgctgattatgattgaaagagaagccttaatt 344
Db 759 KKGKKKKKKKKKKGGGKGGGKAGDDAKDKTKKKKKRAATTTTCKKKKKRAAK 818
Oy 345 tggaaagaagaatcaccgtcattataatcgaaatgattgattgcttgggaag 404
Db 819 KKKAADRTKKTWMDAAAAAATKTKDKGKKKKTKTKKKKKKKKKGGGKKDAAK 878

```

```

Oy 405 ggaatgctggcgtcgaagggtataacaagcaatcccgattgctgaagccggtctaa 464
Db 879 KKKGKKKKKKKKGGGKGGGKGGKGGKGGKDDAAAATKKKKKKKKGGGKGGKGG 938
Oy 465 ccataccttgatgtagatgctgcgggtgggacattccacatcattcatgatatta 524
Db 939 KKGKKKKTKTKTKKKKKKKKRAADAAKGTTKRAADAAAADWTDATTKKKKKDKKT 998
Oy 525 tggaaagaagaatttattgctg 547
Db 999 KTKKKDAAAAATKKKKKKKKKKTKR 1021

RESULT 3
CNS00B3U 1101 bp DNA GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR22N14 of RPCT-98 library from Drosophila melanogaster (fruit
AL056389 genomic survey sequence.
AL056389.1 GI:4937158
GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Location/Qualifiers
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1lb="RPCT-98"
/clone="BACR22N14"
/note="end : TET3"

BASE COUNT 88 a 29 c 122 g 342 t 520 others
ORIGIN
Query Match 4.5%; Score 40.4; DB 121; Length 1101;
Best Local Similarity 22.9%; Pred. No. 0.82;
Matches 55; Conservative 85; Mismatches 100; Indels 0; Gaps 0;

Oy 185 ggaacaaatatacatcacaagaacattatctcttagcgaagaataacatt 244
Db 617 BAGAAMMMSKMTCDTKTBTBTBYADWMTCAATYMTYKYYRAAABDGTYYKMRK 558
Oy 245 tggctgctgcttgaaattcatgattgctaccggtatgattcattgcttcctcaatc 304
Db 557 GMTTRAKCBSTYTYAAKWAATADWMTKAMWKTTTTAATTKMDATATWTKTKTKRA 498
Oy 305 gtgcgttgattatgattgattgaagaagtgcttatttggaaagaagaataaccgtca 364

```

	Db	497	AMKCTATKKKKKKKKGGTAVAKKKAAATAKTATFKKAAAAAWDTATTAA	438
	OY	365	tatatatttggaatgatggttgcctcgsgaaagggagtcgttcgaagggt	424
	Db	437	KAKKKAATKTTAADAAGATGATKTKTAKTAKGAKKKRGDGAAMAATGTGKRTAAKGKK	378
	RESULT	4		
	AQ332958/C		AQ332958	552 bp DNA GSS 06-MAR-1999
	LOCUS		HS.5001.A2.F09.T7.RP.CII Human Male BAC Library Homo sapiens	
	DEFINITION		genomic clone Plate=577 Col=18 Row=K, DNA sequence.	
	ACCESSION		AQ332958	
	VERSION		AQ332958.1 GI:4130585	
	KEYWORDS		GSS.	
	SOURCE		human.	
	ORGANISM		Homo sapIens	
	REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 552)	
	AUTHORS		Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J.J., Zhao,S., Adams,M.D. and Hood,L.	
	TITLE		Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	
	JOURNAL		Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	
	MEDLINE		99380589	
	COMMENT		Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel.: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (Info@resgen.com). BAC end Web Server: http://www.husc.washington.edu Plate: 577 row: K column: 18 Seq primer: T7 Class: BAC ends High quality sequence stop: 552.	
	FEATURES			
	source		Location/Qualifiers	
			1..552	
			/organism="Homo sapIens"	
			/db_xref="taxon:9606"	
			/clone_lib="Plate=577 Col=18 Row=K"	
			/clone_id="RPCII Human Male BAC Library"	
			/sex="Male"	
			/cell_type="Lymphocytes"	
			/note="Vector: pBACe3.6; RPCII Human Male BAC Library"	
	BASE COUNT		177 a 133 c 67 g 170 t 5 others	
	ORIGIN			
	Query Match		4.4%; Score 39.4; DB 96; Length 552;	
	Best Local Similarity		53.7%; Pred.No.1.3;	
	Matches 79; Conservative		0; Mismatches 68; Indels 0; Gaps 0;	
	OY	204	tacagtgaaaccttatcctcttttagcggaagaatacaatttgtctgtctttaagt	263
	Db	182	TAAATAGAATAATCATATATATAGTGATGATGAGAAATATTCCTGCCTCGATGTGATGT	123
	OY	264	tcatgatgccoggttaattgatccattgcttgctgccaatgctgtgtattatgat	323
	Db	122	CCAAGACTGCGAGGTTTTGTGNTTTGAATGACAATGACATGACTCATTCTATGTAATTTNGT	63
	OY	324	tgaatgagaagtgtcttaattgaaa	350
	Db	62	TGATGAGATAGTAAATCAATCGAGGA	36
	RESULT	5		
	AQ621653			

LOCUS	A0621653	520 bp	DNA	GSS	16-JUN-1999
DEFINITION	HS_3089_B2.H12_MF.C1T	Approved Human Genomic Sperm Library	D	Homosapiens	genomic clone Plate=3089 Col=24 Row=P, DNA sequence.
ACCESSION	A0621653				
VERSION	A0621653.1	GI:5084045			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 520) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A.	96 (17), 9739-9744			(1999)
MEDLINE	99380589				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web server: http://www.hsc.washington.edu Plate: 3089 Row: P Column: 24 Seq primer: T7 Class: BAC ends High quality sequence stop: 520. Location/Qualifiers 1..520 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate=3089 Col=24 Row=P" /clone_lib="C1T Approved Human Genomic Sperm Library D" /sex="male" /note="Organ: sperm; Vector: pBelOBAC11; BAC clones in E-Coli DH10B"				
BASE COUNT	146 a	100 c	129 g	143 t	2 others
ORIGIN					
Query Match	4.4%;	Score 39.2;	DB 100;	Length 520;	
Best Local Similarity	57.3%;	Pred. No. 1.5;			
Matches	71;	Conservative 0;	Mismatches 53;	Indels 0;	Gaps 0;
QY	223	tcttaagcggagaataacattgctgctgcttgaagtcacatgatcaccggtat	282		
Db	211	tcttttcgtcgaatcagactctttattttttttttctgttaagtctccacggttt	270		
QY	283	gattccatctgctgcgtcaatcgtgctgtgattatattgattgaagaatgagaagtctta	342		
Db	271	gttttaaatgtgatgtgttcacagatttcggttcacagatttcacactaaagagatgcatttta	330		
OY	343	attg	346		
Db	331	aattg	334		
RESULT	6				
LOCUS	A2138670	750 bp	DNA	GSS	06-JUN-2000
DEFINITION	SP_0170_A2.C04.SP6E	Strongyloidentrotus purpuratus			purple sea urchin, sperm genomic BAC library Strongyloidentrotus purpuratus genomic clone Plate=170 Col=8 Row=E, DNA sequence.
ACCESSION	A2138670				
VERSION	A2138670.1	GI:8290573			
KEYWORDS	GSS.				
SOURCE	Strongyloidentrotus purpuratus.				
ORGANISM	Strongyloidentrotus purpuratus				
	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;				

[illegible][illegible]

Page 9

This Page Blank (uspto)


```

PN W09964619-A2.
XX 16-DEC-1999.
FD 10-JUN-1999; 99WO-DK00314.
XX 10-JUN-1998; 98US-0111256.
XX 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 98DK-0000306.
PR 05-MAR-1999; 98DK-0000307.
PR 05-MAR-1999; 98DK-0000308.
PR 05-MAR-1999; 98DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX (NOVO ) NOVO-NORDISK AS.
XX Kaupinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX WPI: 2000-105891/09.
XX N-PSDB; 245335.
XX New mannanses for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions
XX Claim 1: Page 208-210; 242pp; English.
XX The present sequence represents a mannanase (also known as mannan
XX endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
XX of Bacillus sp. 1633. The mannanase hydrolyses galactomannans.
XX Specifically, mannanses hydrolyse 1,4-beta-D-mannosidic linkages in
XX mannans, galactomannans, glucomannans, and galactoglucomannans. The
XX mannanase protein, or preparations containing it, are used to improve
XX properties of cellulosic or synthetic fibres, yarn or (non)woven
XX fabrics (removal of mannan-based sizes or printing pastes). They are
XX also used to degrade or modify plant materials (particularly recycled
XX waste paper, paper making pulps, or material containing guar or locust
XX bean gums (thickeners), or to reduce viscosity of mannan-containing
XX foods or feeds). The mannanses are also used to process coffee
XX extracts (to inhibit gel formation); in cleaning compositions (for
XX machine washing of fabrics, as hard-surface cleaners, for hand or
XX machine dishwashing, also in oral, dental, contact lens or body-care
XX compositions) where they remove mannan-containing soils and prevent
XX blinding of some soils to cellulotics; and in fabric softeners. They
XX can also be used in oil well drilling to fracture subterranean
XX formations.
XX Sequence 490 AA:
SQ

```

```

Query Match 100.0%; Score 2607; DB 21; Length 490;
Best Local Similarity 100.0%; Pred. No. 1.9e-186;
Matches 490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB 181 naglnhtlmvdaagwqgtpgslnhdygrevlndpqrntmstlmheyaggnasqvrlnid 240
QY 241 RVLNDLALVIGFGRHRTNGDVEATIMSYSEORGVWILAMSKNGPEWEYIDLNDW 300
DB 241 rvlngdlalvlgfgrhrtngdveatimsyseqrgvwlamskngpeweyidlndw 300
QY 301 AGNNLTAWGNTIVNGPYGLRRTSRLSTVFTGGSDGTSPTTLVDFGSMGWTGSSLGC 360
DB 301 agnltawgntivngpyglrretsrstcvftggsgdgtspttlydfegsmgwtgsslg 360
QY 361 GPMVATESSKSGSHKADLOLSSNOHYLVIONTSLOQNSRQATVYKHAHWSVNGM 420
DB 361 gpmvatewssksgshkadtqlssnsqnylvhqnstlqqrstqatvkhahnwsvngm 420
QY 421 TARLYVKHGYTWYSGSFVPINGSSTTLSDLSNQNLSQVREIGVQFQASDSSGQT 480
DB 421 tarlyvktghytwysgsfvpingssttllsdlsnqnlsqvrelsvqfqsadsstg 480
QY 481 STYDNTIVE 490
DB 481 stydnvive:490
RESULT 2
Y44495
ID Y44495 standard; Protein; 493 AA.
AC Y44495;
XX 27-MAR-2000 (first entry)
DE Bacillus agaradherens NCIMB 40482, Mannanase enzyme.
XX Mannanase: endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
XX detergent composition; mid-branched anionic surfactant; washing;
XX cosmetic stain; food stain.
XX OS Bacillus agaradherens.
XX FH Key
XX FT Peptide 1..32
XX FT Protein 33..343
XX FT /label= Mature_Mannanase
PN W09964552-A1.
XX 16-DEC-1999.
XX 10-JUN-1998; 98WO-US12026.
XX 10-JUN-1998; 98WO-US12026.
XX 10-JUN-1998; 98WO-US12026.
XX (PROC ) PROCTER & GAMBLE CO.
XX PA Bettloll JP, Thoen CAJK;
XX PI WPI: 2000-116536/10.
XX DR N-PSDB; Z29845.
XX Detergent composition for removing greasy stains such as cosmetics,
XX food stains and body soils
XX PT food stains and body soils
XX PS Disclosure; Page 102; 113pp; English.
XX The present sequence is the B. agaradherens NCIMB 40482, alkaline
XX mannanase enzyme. Mannanase shows maximum activity at pH ranging
XX from 7.5-10.5. It can be used in a detergent composition along with a
XX mid-branched anionic surfactant. The detergent composition may be used
XX for washing purposes, soaking/pretreatment of stained fabric, hard
XX surface cleaning and for removal of cosmetic and/or food stains. This
XX composition provides excellent cleaning effect at low temperature.

```



```

OY 251 IGEFGRHRTNGDVDEATIMSYSEORGVGLAMSKNGPEWEYLDLSNDVAGNNTLTMGN 310
DB 251 Igefghrhtcdgdvdedtlisyeetgvlawskngstewdyldlsedwagqhltdwgn 310
OY 311 TIVNGPYGLRETSRLSTVFT--GGSGDGTSPPTLYDFEGSMGWTGSSLSGGPWAVTEW 368
DB 311 rlvbgadglgetskpvtvfdndngphpeptatcltydfegstgwhsnvtygspwvlew 370
OY 369 SSKGSHSLKADLOLSNSQHTLVIONTSIQONSRIQATVHANMWSYVNGMTARLYVKT 428
DB 371 gasgnyalkadvnltsnshelysegrnlhysqqlnatvrvhanvgnpgumariyykt 430
OY 429 GHGYTWYSGSFVPINGS-SGTTLSIDLNSVQNTLSQVREIGVQFSASDSSGQISITIDNV 487
DB 431 gsdtywhsgpfrlnssngstltsfdlnlnshvrelgyvfasadnsqgatalydvhn 490

RESULT 4
Y44496
ID Y44496 standard; Protein; 468 AA.
AC Y44496;
DE 27-MAR-2000 (first entry)
XX Bacillus agaradherens Clone MB594, Mannanase enzyme.
XX
XX Mannanase: endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KW detergent composition; mid-branched anionic surfactant; washing;
KM cosmetic stain; food stain.
OS
XX Bacillus agaradherens.
XX Synthetic.
FH Key
FT Peptide 1, 31
FT Protein /label- Signal_peptide
FT /label- Mature_Mannanase
PN WO9964552-A1.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Bettiol JP, Thoen CAWK;
XX
XX WPI; 2000-116536/10.
XX
XX N-PSDB; 229846.
XX
XX Detergent composition for removing greasy stains such as cosmetics,
PT food stains and body soils
XX
XX Disclosure; Page 103-104; 113pp; English.
XX
XX The present sequence is the B. agaradherens Clone MB594, alkaline
XX Mannanase enzyme. Clone MB594 is derived from B. agaradherens strain
XX NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
XX from 7.5-10.5. It can be used in a detergent composition along with a
XX mid-branched anionic surfactant. The detergent composition may be used
XX for washing purposes, soaking/pre-treatment of stained fabric, hard
XX surface cleaning and for removal of cosmetic and/or food stains. This
XX composition provides excellent cleaning effect at low temperature.
XX
XX Sequence 468 AA;
SQ

```

Query Match 65.6%; Score 1711; DB 21; Length 468;

```

Best Local Similarity 69.3%; Pred. No. 9,2e-120;
Matches 314; Conservative 60; Mismatches 75; Indels 4; Gaps 3;
OY 11 ITSLILASSLIFVSGSTPAMNSGFFVSGTLYDANGNPVYMRGINHGANYKQOATTA 70
DB 12 lltclltlsvynglgt-tspsaastgfyvngncltydangqpfvmrglnbhavykdtasta 70
OY 71 IESANTGANTVAVIVSDGGQRTKDIHVRNLISLEDNHLYVAPEVDATGYOSIASL 130
DB 71 lpaiaegantllrlvlisdgqwekdldtrelaelaegnmvavevndaqgrdscl 130
OY 131 NRVADYIEMRSALICEKEDVINININEMWGSWEGDAMADYKQALPRLNGLNHTLMV 190
DB 131 nrvadvylemkdldlkkedvrlinlanewyswdsawadgyldvlpklrdagltchlmv 190
OY 191 DAAGWGFQSIHDYGEVFNADPQRNMFSTIMYIYAGNSQVNTNIDRYLNOLALV 250
DB 191 daagwgfypsihdygqvfndpqlkntmflmyeyaggdantvrsndrvldqdlav 250
OY 251 IGEFGRHRTNGDVDEATIMSYSEORGVGLAMSKNGPEWEYLDLSNDVAGNNTLTMGN 310
DB 251 Igefghrhtcdgdvdedtlisyeetgvlawskngstewdyldlsedwagqhltdwgn 310
OY 311 TIVNGPYGLRETSRLSTVFT--GGSGDGTSPPTLYDFEGSMGWTGSSLSGGPWAVTEW 368
DB 311 rlvbgadglgetskpvtvfdndngphpeptatcltydfegstgwhsnvtygspwvlew 370
OY 369 SSKGSHSLKADLOLSNSQHTLVIONTSIQONSRIQATVHANMWSYVNGMTARLYVKT 428
DB 371 gasgnyalkadvnltsnshelysegrnlhysqqlnatvrvhanvgnpgumariyykt 430
OY 429 GHGYTWYSGSFVPINGS-SGTTLSIDLNSVQNTLSQVREIGVQFSASDSSGQISITIDNV 487
DB 431 gsdtywhsgpfrlnssngstltsfdlnlnshvrelgyvfasadnsqgatalydvhn 490

RESULT 5
Y54125
ID Y54125 standard; Protein; 468 AA.
AC Y54125;
DE 27-MAR-2000 (first entry)
XX
XX Amino acid sequence of a Bacillus mannanase enzyme.
XX
XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW mannan; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture.
XX
XX Synthetic.
XX Bacillus agaradherens.
XX
XX WO9964619-A2.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1999; 99WO-DK00314.
XX
XX 10-JUN-1998; 98US-0111256.
XX
XX 20-OCT-1998; 98DK-0001340.
XX
XX 20-OCT-1998; 98DK-0001341.
XX
XX 28-OCT-1998; 98US-0105970.
XX
XX 28-OCT-1998; 98US-0106054.
XX
XX 23-DEC-1998; 98DK-0001725.
XX
XX 05-MAR-1999; 98DK-0000306.
XX
XX 05-MAR-1999; 99DK-0000307.

```

[illegible]

Db	371	gasgnylskadvnltstnseshelyegsrnlngysqlnatvtrhanwmpngmarlyvkt	430
Oy	429	GHGYTWSSGFVPINGS-SGTTLSIDLSNVQL	460
Db	431	gsdytwhsnpfltrlnssngtllsdlmnenl	463
RESULT	6		
ID	YS4123	standard; Protein; 476 AA.	
XX	YS4123;		
XX	27-MAR-2000	(first entry)	
DE	A mannanase-linker-cellulose binding domain fusion protein.		
XX			
KW	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;		
KW	endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;		
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;		
KW	galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;		
KW	printing paste; plant material degradation; recycled waste paper;		
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;		
KW	mannan-containing food; coffee extract; cleaning composition;		
KW	machine washing; hard-surface cleaner; dishwashing; oral; dental;		
KW	contact lens; body-care composition; fabric softener; oil well drilling		
KW	subterranean formation fracture; cellulose binding domain.		
XX			
OS	Synthetic.		
OS	Bacillus sp.		
OS	Clostridium thermocellum.		
XX			
XX	WO9964619-A2.		
PD	16-DEC-1999.		
XX			
PF	10-JUN-1999;	99WO-DK00314.	
XX			
PR	10-JUN-1998;	98US-0111256.	
PR	20-OCT-1998;	98DK-0001340.	
PR	20-OCT-1998;	98DK-0001341.	
PR	28-OCT-1998;	98US-0105970.	
PR	28-OCT-1998;	98US-0106054.	
PR	23-DEC-1998;	98DK-0001725.	
PR	05-MAR-1999;	99DK-0000306.	
PR	05-MAR-1999;	99DK-0000307.	
PR	05-MAR-1999;	99DK-0000308.	
PR	05-MAR-1999;	99DK-0000309.	
PR	09-MAR-1999;	99US-0123543.	
PR	10-MAR-1999;	99US-0123623.	
PR	10-MAR-1999;	99US-0123641.	
PR	11-MAR-1999;	99US-0123642.	
PA	(NOVO) NOVO-NORDISK AS.		
XX			
PI	Kauppinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;		
DR	WPI; 2000-105891/09.		
DR	N-PSDB; 245336.		
XX			
PT	New mannanases for treatment of textiles, plant material and coffee		
PT	extract, and in cleaning compositions		
XX			
PS	Example 4; Page 211-212; 242pp; English.		
CC	The present sequence represents a mannanase-linker-cellulose binding		
CC	domain fusion protein. Mannanase (also known as mannan		
CC	endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)		
CC	hydrolyses galactomannans. Specifically, mannanases hydrolyse		
CC	1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,		
CC	and galactoglucomannans. The mannanase protein, or preparations		
CC	containing it, are used to improve properties of cellulosic or		

Db	11	alic111fagvlnltssgqeahnsgfhvngttcllydangnprfcmvrlnglnhgwfkxleqt	70
Qy	70	AIEGIANTGANTVRIIVLSDGGQWTKDDIHTVRNLISLAEDNHLVAVPEVHATGIDSIAS	129
Db	71	smrsgsqgqantllivlsnsgwqkddrmnvavslslaeghqmlavlevhdatsnfnad	130
Qy	130	LNRADVIEIEMRSLLICKEDFVIITINEMFGSMEGMADGKQALPRLRNAGLHRTLM	189
Db	131	lgaevdyvleukdvlygqfedvlinlnenewygawdggawerqymalrlqrlnaglshtlm	190
Qy	190	VDAAGWGFQPSIHDIHQREYFNADPQNTMFESIHMYEYAGNAGNSQVRTNIDRYLNODLL	249
Db	191	vdaagygypspsvdydgyevlneapqntmfsvhmyeyagdaenrttrnldslsngla	250
Qy	250	VIEGFGRHRTNGVDDEATIMSYSEQRCVGLMWSKNGPWEYETLDLSNDMAGNNLTWAG	309
Db	251	vlgfegrhvlygqdvdedtlllsyqgrnvglawsvbngnsegveyldlndfagnrltwg	310
Qy	310	NTIYNGPYGLAETSRSLSTVF	329
Db	311	drlyvngpyglrqrskrsrvf	330
RESULT	8		
Y54129			
ID	Y54129	standard; Protein; 369 AA.	
XX	Y54129;		
DT	27-MAR-2000	(first entry)	
DE		Amino acid sequence of a <i>Bacillus</i> sp. Aa349 mannanase enzyme.	
XX			
XX		Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;	
KW	endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;		
KW	mannan; galactomannan; glucomannan; galactoglucomannan; cellulose fibre		
KW	synthetic fibre; yarn; fabric; printing paste; chicken; viscosity;		
KW	plant material degradation; recycled waste paper; paper making pulp;		
KW	guar; locust bean gum; mannan-containing food; coffee extract;		
KW	cleaning composition; machine washing; hard-surface cleaner;		
KW	dishwashing; oral; dental; contact lens; body-care composition;		
KW	fabric softener; oil well drilling; subterranean formation fracture.		
XX			
OS	Bacillus sp.		
XX			
XX	Key	Location/Qualifiers	
FT	Peptide	1..25	
FT	Domain	/note- "signal peptide"	
FT	Domain	26..67	
FT	Domain	/note- "N-terminal domain of unknown function"	
FT	Domain	68..369	
XX		/note- "catalytic domain"	
PN	W09964619-A2.		
XX			
PD	16-DEC-1999.		
XX			
PF	10-JUN-1999;	99WO-DK00314.	
XX			
PR	10-JUN-1998;	98US-0111256.	
PR	20-OCT-1998;	98DK-0001340.	
PR	20-OCT-1998;	98DK-0001341.	
PR	28-OCT-1998;	98US-0105970.	
PR	28-OCT-1998;	98US-0106054.	
PR	23-DEC-1998;	98DK-0001725.	
PR	05-MAR-1999;	99DK-0000306.	
PR	05-MAR-1999;	99DK-0000307.	
PR	05-MAR-1999;	99DK-0000308.	
PR	05-MAR-1999;	99DK-0000309.	
PR	09-MAR-1999;	99US-0123543.	
PR	10-MAR-1999;	99US-0123623.	
PR	10-MAR-1999;	99US-0123641.	
PR	11-MAR-1999;	99US-0123642.	

[illegible]

XX	synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;	
KW	plant material degradation; recycled waste paper; paper making pulp;	
KW	guar; locust bean gum; mannan-containing food; coffee extract;	
KW	cleaning composition; machine washing; hard-surface cleaner;	
KW	dishwashing; oral; dental; contact lens; body-care composition;	
KW	fabric softener; oil well drilling; subterranean formation fracture.	
XX		
OS	Bacillus sp.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..28
FT		/note- "signal peptide"
FT	Domain	29..320
FT		/note- "catalytic domain"
XX		
PN	W09964619-A2.	
PD	16-DEC-1999.	
XX		
PF	10-JUN-1999;	99WO-DK00314.
XX		
PR	10-JUN-1998;	98US-0111256.
PR	20-OCT-1998;	98DK-0001340.
PR	20-OCT-1998;	98DK-0001341.
PR	28-OCT-1998;	98US-0105970.
PR	28-OCT-1998;	98US-0106054.
PR	23-DEC-1998;	98DK-0001125.
PR	05-MAR-1999;	99DK-0000306.
PR	05-MAR-1999;	99DK-0000307.
PR	05-MAR-1999;	99DK-0000308.
PR	05-MAR-1999;	99DK-0000309.
PR	09-MAR-1999;	99US-0123543.
PR	10-MAR-1999;	99US-0123623.
PR	11-MAR-1999;	99US-0123641.
PR	11-MAR-1999;	99US-0123642.
PA	(NOVO) NOVO-NORDISK AS.	
XX		
FI	Kauppinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;	
DR	WPI: 2000-105891/09.	
DR	N-PSDB: 245345.	
XX		
PT	New mannanases for treatment of textiles, plant material and coffee	
PT	extract, and in cleaning compositions	
XX		
PS	Claim 35; Page 228-229; 242pp; English.	
XX		
CC	The present sequence represents a mannanase enzyme (also known as	
CC	mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).	
CC	The mannanase hydrolyses galactomannans. Specifically, mannanases	
CC	hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,	
CC	glucomannans, and galactoglucomannans. The mannanase protein, or	
CC	preparations containing it, are used to improve properties of cellulosic	
CC	or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based	
CC	sizes or printing pastes). They are also used to degrade or modify	
CC	plant materials (particularly recycled waste paper, paper making pulps,	
CC	or material containing guar or locust bean gums (thickeners), or to	
CC	reduce viscosity of mannan-containing foods or feeds). The mannanases	
CC	are also used to process coffee extracts (to inhibit gel formation); in	
CC	cleaning compositions (for machine washing of fabrics, as hard-surface	
CC	cleaners, for hand or machine dishwashing, also in oral, dental, contact	
CC	lens or body-care compositions) where they remove mannan-containing	
CC	soils and prevent binding of some soils to cellulosics; and in fabric	
CC	softeners. They can also be used in oil well drilling to fracture	
CC	subterranean formations.	
XX		
XX	Sequence 320 AA:	

Query Match	36.78;	Score 957.5;	DB 21;	Length 320;
Best Local Similarity	58.48;	Pred. No. 7.1e-64;		
Matches 181;	Conservative 42;	Mismatches 86;	Indels 1;	Gaps 1

Accession	Protein Name	Length (aa)	Source
QY16	LLASLIFVSGISTANANSGFYVSGTTLTYLDANGNPFVNRGINHGHAAMKKDQATTAISGIA	75	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
Db12	lvavlvlfvysglaaagsgfhwkgfelleldknngdpyvmrghvnhawtkqgleeipala	71	endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
QY76	NTGANVRIYLSGGGGWMTKDDIHVTNLTSLAEDNHLVAPEVHDATGSDSIASINRAVD	135	mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic
Db72	eegenvtrvrlsnsgqgwekdaseelarvlaatecylgtltvlevnhatsgsdnpddkavd	131	synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
QY136	YVIEIMRSALIGKEDFYIINIANFWGSGWEGDAMADGYKQALPRLNAGLHNTLWADAAG	195	plant material degradation; recycled waste paper; paper making pulp;
Db132	lyvlemaadvllygtedrvlllnaenewgawswsdwaeaqaprlrlsaglahtllvydaegw	191	gum; locust bean gum; mannan-containing food; coffee extract;
QY196	GQFQPSIHDGGRVFNADPQRNMFSEIHMYEYAGAGNASQVRNIDRVLNODLATYISGEG	255	cleaning composition; machine washing; hard-surface cleaner;
Db192	qyypaslhhegadvfasdpdklkmfmslhmeyagadarvseenldgvaenlavvlgfeg	251	dishwashing; oral; dental; contact lens; body-care composition;
QY256	HRHNGVDVDAITMSYSEQRCYVGLMWSKGVGPEWEYLDLSNDWAGNNLTAMGNTIYNG	315	fabric softener; oil well drilling; subterranean formation fracture.
Db252	hrhdgdvddedallaylaerqyvwslawswyngsgvyejldleqpsg-plawgerlvyy	310	
QY316	PYGLRETSRL325		
Db311	emgllkvldhl320		
RESULT	10		
Y54130	Y54130 standard; Protein: 305 AA.		
XX	Y54130;		
XX	27-MAR-2000 (first entry)		
DE	Amino acid sequence of a partial Bacillus sp. mannanase enzyme.		
XX			
XX	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;		
KW	endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;		
KW	mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic		
KW	synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;		
KW	plant material degradation; recycled waste paper; paper making pulp;		
KW	gum; locust bean gum; mannan-containing food; coffee extract;		
KW	cleaning composition; machine washing; hard-surface cleaner;		
KW	dishwashing; oral; dental; contact lens; body-care composition;		
KW	fabric softener; oil well drilling; subterranean formation fracture.		
XX			
OS	Bacillus sp.		
XX			
PN	W09964619-A2.		
XX			
PD	16-DEC-1999.		
XX			
PE	10-JUN-1999; 99MO-DK00314.		
XX			
XX	10-JUN-1998; 98US-0111256.		
PR	20-OCT-1998; 98DK-0001340.		
PR	20-OCT-1998; 98DK-0001341.		
PR	28-OCT-1998; 98US-0105970.		
PR	28-OCT-1998; 98US-0106054.		
PR	23-DEC-1998; 98DK-0001725.		
PR	05-MAR-1999; 99DK-0000306.		
PR	05-MAR-1999; 99DK-0000307.		
PR	05-MAR-1999; 99DK-0000308.		
PR	05-MAR-1999; 99DK-0000309.		
PR	09-MAR-1999; 99US-0123543.		
PR	10-MAR-1999; 99US-0123623.		
PR	10-MAR-1999; 99US-0123641.		
PR	11-MAR-1999; 99US-0123642.		
XX			
PA	(NOVO) NOVO-NORDISK AS.		
XX			
PI	Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;		
XX			

CC product into vector pDN1981. DNA encoding the cellulose binding

2
1
1
6
9
6
3
2
4
3
1

CC product into vector pDN1981. DNA encoding the cellulose binding

XX WPI: 1997-402610/37.
DR N-PSDB: T77055.

PT Desizing cellulose⁶-containing fabric or textile using an enzyme
PR hybrid - which comprises a catalytically active amino acid sequence
PT of a non-cellulolytic enzyme linked to an amino acid sequence
PT comprising a cellulose binding domain.

XX Example 2; Page 52-54; 72pp; English.

PS This protein comprises the alkaline cellulase (endoglucanase) of
CC Bacillus agaradherens NCIMB 40482. It was expressed in Bacillus
CC subtilis PL2306 transformants following PCR amplification (see
CC T77053-54) of B. agaradherens genomic DNA and ligation of the PCR
CC product into vector pDN1981. DNA encoding the cellulose binding
CC domain (CBD) of the alkaline cellulase was subsequently amplified
CC (see T77056-57) for use in the construction of a novel alpha-
CC amylase-CBD hybrid enzyme (see T77058). A claimed process for
CC desizing cellulose-containing fabric or textile comprises treating
CC the fabric or textile with a modified enzyme (enzyme hybrid)
CC comprising a non-cellulolytic enzyme linked to a CBD. The process
CC gives improved enzyme performance by modifying the enzyme so as to
CC increase its affinity for cellulosic fabric.

XX Sequence 400 AA;

Query Match 8.8%; Score 229.5; DB 18; Length 400;
Best Local Similarity 23.4%; Pred. No. 1.5e-09;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

QY 6 KRFEFTLSLLASSTLFSGSTANANS-----GPFVSGTLYDANGNPFVNRGI-NH 58
DB 2 KRtctlfvvlmvtvalfsygnltaadsvveengqlslsngelvnergeqvqlkqmsn 61
QY 59 GRAMV-----KQDATTAEIGANTGANTVRIVL--SDGGQWTKDI-HTVRNLT 103
DB 62 glqwygqfvnyesmkwlr-----dwglhvfaamytsagyliddpsvkekyka 111
QY 104 ISIAEDNHLVAVPEVDATGYDSIASLNRAVDYWIEMKRSALIGKEDTVITINANWFGSW 163
DB 112 veealldldyvlidwhlslsdnpnylkeekdfidem-selygdpnylyelanepgs- 169
QY 164 EGDAMADGVR---QAIPRLRNAGLWHTLWDAAGQGQFPOSIHGGRVFNADPQRNTW 219
DB 170 -dvtwgnqlkpyaeewipilrindpnmlitv---gctwsqdvh-haadnqladp--nvm 222
QY 220 FSIHMYEYAGNAGSOVRTNIDRYLNODLALVIGFGRHRTNGD---VDEATI-MSYSEQ 274
DB 223 yafhf--yagthgnrlrdqvdaldgaaalfvegwtsaatggyvfldeagwldmde 280
QY 275 RGYGMLAWS-----WKGNGPMEWELDLSNDMAGNNLTMGNTIVNGPYGLRETS 323
DB 281 rnlswanwsltlhdeesaaalmpgarp-----tgywteaelpsgflvre--klres 330
QY 324 RLTSTVFTGGSDGSGTPTTLYDFEGSMQWGTGSSL-----SGGPWATWEWS 369
DB 331 slppsdptppsdgppdtpcpdpggyapawdpqilyrnelvymgqlwqakwt 384

Search completed: December 19, 2000, 16:49:31
Job time: 1181 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:41:28 ; Search time 156.76 Seconds
(Without alignments)
52.392 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LINGFKKISITSLILASS.....QASDSSGQTSTIDNYIV 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCtUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/Backfil1est.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229.5	8.8	400	2	US-08-713-298B-2
2	229.5	8.8	400	2	US-08-870-180B-2
3	229.5	8.8	400	3	US-08-814-052-4
4	229.5	8.8	400	3	US-08-812-829-4
5	229.5	8.8	462	2	US-08-870-180B-13
6	213	8.2	467	2	US-08-727-548-2
7	189	7.2	551	2	US-09-033-537A-1
8	174	6.1	430	2	US-08-924-440-2
9	160	6.1	531	2	US-07-862-588B-7
10	143	5.5	521	1	US-08-276-213-3
11	142	5.4	358	1	US-08-604-913B-11
12	129	4.9	1536	1	US-08-038-662-2
13	129	4.9	1536	1	US-08-302-832-2
14	129	4.9	1536	2	US-08-530-198-2
15	129	4.9	1536	2	US-08-469-880-2
16	129	4.9	1536	2	US-08-728-470-2
17	129	4.9	1536	2	US-08-617-697-2
18	123.5	4.7	880	3	US-08-947-965-77
19	116.5	4.5	983	3	US-08-164-292B-26
20	116.5	4.5	983	3	US-08-845-623-26
21	116.5	4.5	983	3	US-08-815-927-26
22	112.5	4.3	490	1	US-08-672-571A-1
23	111	4.3	574	3	US-08-732-433-1
24	109	4.2	735	2	US-08-313-185-48
25	109	4.2	735	2	US-08-459-489-9
26	108	4.1	1545	4	PCT-US95-10661A-4
27	107.5	4.1	666	2	US-08-737-716-14
28	107.5	4.1	1222	2	US-08-682-517-15

29	107.5	4.1	1252	2	US-08-682-517-9	Sequence 9, Appl
30	106.5	4.1	750	5	5457037-3	Patent No. 5457037
31	106	4.1	1338	2	US-08-728-470-9	Sequence 9, Appl
32	106	4.1	1599	2	US-08-617-697-9	Sequence 9, Appl
33	105.5	4.0	751	5	5457037-5	Patent No. 5457037
34	105	4.0	385	2	US-08-387-942C-26	Sequence 26, Appl
35	105	4.0	654	1	US-08-392-828C-2	Sequence 2, Appl
36	105	4.0	654	3	US-09-330-945-2	Sequence 4, Appl
37	105	4.0	997	2	US-08-387-942C-4	Sequence 4, Appl
38	104	4.0	398	2	US-08-853-659A-45	Sequence 12, Appl
39	104	4.0	735	2	US-08-459-499-12	Sequence 12, Appl
40	103.5	4.0	553	2	US-08-661-052-16	Sequence 15, Appl
41	103	4.0	434	2	US-08-989-925-1	Sequence 1, Appl
42	103	4.0	1702	4	PCT-US95-10661A-5	Sequence 5, Appl
43	102.5	3.9	623	3	US-09-041-991A-6	Sequence 6, Appl
44	101.5	3.9	566	2	US-07-862-588B-4	Sequence 4, Appl
45	101.5	3.9	928	1	US-08-474-140-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-713-298B-2
Sequence 2, Application US/08713298B
Patent No. 5922586
GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dammann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Sch lein, Martin
APPLICANT: J rgensen, Per
APPLICANT: Bjoernvad, Mads
TITLE OF INVENTION: DNA Constructs and Methods of Producing
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59225860 No. 5922586disk of No. 5922586th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,298B
FILING DATE: 13-SEPT-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3794,424-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 867-0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-298B-2

Query Match 8.8%; Score 229.5; DB 2; Length 400;
Best Local Similarity 23.4%; Pred. No. 5.8e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

```

01  TS-08-870-180B-2
02
03  Query Match      8.8%; Score 229.5; DB 2; Length 400;
04  Best Local Similarity 23.4%; Pred. No. 5.8e-12;
05  Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19
06
07  QY      6 KIIISITLSLASSILIFSGSTANANS-----GFVSGTLLYDANGNEFYMRGI-NH 58
08      2 KIIITIVVLLMVALFISGNTTAADNDSVVEHGQLSISNGELVNERGEQVQLKGMSSH 61
09
10  QY      59 GHAVY-----KQDATTALBGIANTGANTVRVYL--SDGGCKTKNDI-HYRNL 103
11      62 GLQWYGFVNEYSKMYLRD-----DWGINFRAMTSSGGYIDDSVEKEKVEA 111
12
13  QY      104 ISLAEDNHVAAPVEVDATGATGDSIASLNAVDVYIEMRSALIGKEDVYIIANENFGSW 163
14      112 VEAADIDDIYIIDMHLISDNDEPNITKEAKKFFDEDM-SELXGDVPNTIYIEADNPAGNS- 169
15
16  QY      164 EGDAMADGVR---QAIPRLRNAGLNHTLMVDADAAGMGOFPOSINDHYGREVENADPQRNTM 219
17      170 -DVTWNGNQIPAEVIEVPIIRNNDPNNIIIV--GIGTWSQDVH-HAADNQLADP--NVM 222
18
19  QY      220 FSIHMYETAGAGNSQVFTNIDRVLLNODLALVIGEGFHRITND-----VDENI-MSYSEQ 274
20      223 YAFHF--YAGHGGONLDDQVDYALADGGAEIFVSEWTSANTDGGVFLIDAEQWVWIDFENDE 280
21
22  QY      323 RGYGMIAWS-----WRNGGEWEYELDSNDAGNNTIANGNTIVAGPYGLRETS 323

```

324 RLSTVEYGGGGGDTSP7LLYDFEGSMQGTGSL-----SCGPAAVTEWS 369
 331 SIPSPDPTPPDDPGEPPDP7PPSPDGEXEYAMPDPQIYNEIYVHNGOLQWAKMT 384
 RESULT 3
 US-08-814-052-4
 Sequence 4, Application US/08814052
 Patent No. 6015783
 GENERAL INFORMATION:
 APPLICANT: von der Osten, Claus
 APPLICANT: Chertiy, Joel R.
 APPLICANT: Bjornvad, Mads E.
 APPLICANT: Vind, Jesper
 APPLICANT: Rasmussen, Michael Dolberg
 TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
 TITLE OF INVENTION: OR STAINS FROM CELLULOISIC FABRIC
 NUMBER OF SEQUENCES: 55
 CORRESPONDENCE ADDRESS:
 ADDRESSSEE: No. 6015783 No. 6015783disk of No. 6015783th America, Inc
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastISO for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/814,052
 FILING DATE: 06-MAR-1997
 CLASSIFICATION: 510
 ATTORNEY/AGENT INFORMATION:
 NAME: Lambiris, Elias J
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 4684, 204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:

RESULT 4
US-08-812-829-4
: Sequence 4, Application US/08812829
: Patent No. 6017751
: GENERAL INFORMATION:
: APPLICANT: von der Osten, Claus
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Wind, Jesper
: APPLICANT: Rasmussen, Michael Dolberg
: TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
: TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 60177510 No. 6017751disk of No. 6017751th America, Inc
: STREET: 405 Lexington Avenue, Suite 6400
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSO for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/812,829
: FILING DATE: 06-MAR-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambiris, Elias J

```

0Y      6 KIIIFITSLILASILEVSTSTANNAS-----GEFVSTTYLYDNGNPEVRG-I-NH 58
Db      2 KIIITIFVYILMTALVBSIGITTAADNDVSVEEHOGLSISNGELVNERGEVOLAKGSSH 61
0Y      59 GHAWT-----KDOATTALEGIAHTGANTVRIVL--SDGGOWTKDDI-HTVBNTL 103
Db      62 GLOWGQEVNYESMKWLRD-----DWGINVFRAAMTSSGGYIDDPVKEYEKEA 111
0Y      104 ISLADNHLVAVPEVHDATGYDTSIASLNRANDVYIEIKRSALIGEDYIYINANENWGSW 165
Db      112 YEAAALDIDYIYIDWHILISDDMDPIYKEEAADFDEM-SELXYGYPNVIYEAINEPGRS- 165
0Y      164 EGDAAWDGYSK-----QAPRLNAGLNHTLMYDAGMCOFPOSIDHYGREYFNADPQRTM 219
Db      170 -DYTWGNQIKRYAEVPIPIINNDPNNTIIV---GTGWSQDVH-HAADNOLADP--NVM 222
0Y      220 FSIHMYEYAGGNASOVRTNIDRVILNODLALVIGEFGRHTNGD---VDEATTI-MSYSEQ 274
Db      223 YAFHF--YAGTHGQULRQOVYDALDOGAALIVSEMGTSAAATGDCGVFLYDEQAVWIDFME 280
0Y      275 RGVGHTLANS-----WKNGCFEVEYILDLSDMAAGNLTLMGWNTIYNGPGLAETS 332
Db      281 RNLSTAMNLSLTHKDESSAALMPGANP-----TGGWTEALELSPSGCFVYEB--KIRESA 330
0Y      324 RLSTVTFGGGSDGGSPTTLDFDFSGMQGWTGSSL-----SGGPAWATLMS 369
Db      331 SIPSPDPPSPSDPGEBPDPPTSPSDGCEETPANDPNQIYINELIYHNGQLQAKWMT 384

```

RESULT 5
 US-08-870-180B-13
 Sequence 13. Application US/08870180B
 Patent No. 5945327.
 GENERAL INFORMATION:
 APPLICANT: Outtrup, Helle
 APPLICANT: Damdbmann, Claus
 APPLICANT: Olsen, Arne
 APPLICANT: Bisgard-Frantzen, Henrik
 APPLICANT: Schleich, Martin
 APPLICANT: Jrgensen, Per
 APPLICANT: Bjornvad, Mads
 TITLE OF INVENTION: DNA Constructs and Methods of Producing
 TITLE OF INVENTION: Cellulytic Enzymes
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 59453270 No. 5945327/disk of No. 5945327th America, Inc
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,180B
FILING DATE: 6-June-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794,434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-870-180B-13

```

```

Query Match      8.8%; Score 229.5; DB 2; Length 462;
Best Local Similarity 23.4%; Pred. No. 7.3e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

```

```

QY 6 KRTFSTLSLLASSILFVSGTSTANANS-----GFVSGTLYDANGNPFVNRGI-NH 58
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 KRTTTFVLLMTALFSGTNTAADNDVVEEHQSLISNGELVNERGEQVLKGMSSH 61
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 GHAWY-----KQDQATAIEGIANTGANTVRIYL--SDGGQWTKDDI-HYVRNL 103
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 GLQMTGTFVYESKRWLKD-----DWGIVFPAAMTSSGGYIDDPVYKEKYEKA 111
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 ISLAEDNHLVAPEVHDATGYDSIASLNRAVDYIEKRSALIGKEDTVIINANEMFSW 163
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 112 VEAALDIDVILDMHILSDNDPNIYKEAKDFEDM--SELYGDPYVPIVEINERPGS- 169
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 EGAMADGIR-----QAIPRLRNAGLNHTLAVDAAGKGFQPSIHIDYREVFNAIDPQARNTY 219
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170 -DVTWGQIPIYAEVPIPIIRNDPNNIIV--GTGTWSQDVH-HAADQLADP--NVW 222
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 220 FSIHMYEAGNNAOVRTNIDRVLNODLALVIGFGRHTNGD---VDEATI--MSYSQ 274
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 223 YAHF--YAGTHGQNLRDQVDYALDGAALFVSEMGISATGSGVFLDERQVWIDPFMD 280
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 RGVGWILWS-----WKGNGPEWEYLDLSNDWAGNNLTAMGNTIVNGPYGLRETS 323
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 281 RNLNMANWSLTHKDESSALMPGANP-----TGWTAEALSPSGTFVRE--KIRESA 330
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 RLSTVFTGGSSDGTSTTLYDEGSGMGWGSSL-----SGGWAIVTEWS 369
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 331 STPPSDPTPPSDGEPDPTPPSDPGKYPAMDPOIYTNEIYVHNGQLMOAKMWT 384
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
US-08-727-548-2
Sequence 2, Application US/08727548
Patent No. 5856165

```

```

GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: No. 5856165el Alkaline Cellulase and Methods of
TITLE OF INVENTION: Producing Same
NUMBER OF SEQUENCES: 2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,548
FILING DATE:

```

```

CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-727-548-2

```

```

Query Match      8.2%; Score 213; DB 2; Length 467;
Best Local Similarity 22.8%; Pred. No. 2e-10;
Matches 99; Conservative 62; Mismatches 157; Indels 116; Gaps 22;

```

```

QY 6 KRTFSTLSLLASSILFVSGTSTA-----NANGFVSGTLYDANGNPFVNRGI-N 57
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 KRTTTFVLLMTALFSGTNTAADNDVVEEHQSLISNGELVNERGEQVLKGMSSH 60
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 GHAWYKQDQATAIEGIANTGANTVRIYLSDDGQWTKDD-----IHT----- 99
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 HGLQMTG-----OFVYESKRWLKDMDGILTFPAAMTSSGGYIDDP 102
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 100 -----VRNLISLAEDNHLVAPEVHDATGYDSIASLNRAVDYIEKRSALIGKEDTVIIN 154
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 103 SVYKEKVEVEAALDGIYVILDMHILSDNDPNIYKEAKDFEDM--SELYGDPYVPIVE 161
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 IANEMFGS--WEGD--ANADGKQALPRLRNAGLNHTLAVDAAGKGFQPSIHIDYGREY 209
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 IANEPGSDVTDMDQIKPFAE--EVIPIVRNDPNNIIV--GTGTWSQDVH-HAADN 214
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 FNADPQRTNFSIHMYEAGNNAOVRTNIDRVLNODLALVIGFGRHTNGD---VDE 265
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 215 QLADP--NMYIAHF--YAGTHGQNLRDQVDYALDGAALFVSEMGISATGSGVFLDE 270
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 ARI-MSYSQGVGWILWS-----WKGNGPEWEYLDLSNDWAGNNLTAMGNTIV 313
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 271 AQWVIDFMDERNLNMANWSLTHKDESSALMPGANP-----TGWTAEALSPSGTFV- 322
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 NGPYGLRETSKLSVY-----TGGSDGSGTSTTLYDEGSGMGWGSSL----- 358
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 323 -----REKIRISASITPPSDPTPPSDGEPDPTPPSD--PGKYPAMDNDPOIYTNEI 375
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 ---SGGWAIVTEWS 369
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 376 VYHNGQLMOAKMWT 389
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
US-09-033-537A-1
Sequence 1, Application US/09033537A
Patent No. 5958083

```

```

GENERAL INFORMATION:
APPLICANT: Onishi, Masahiro
APPLICANT: Fich, Merete
APPLICANT: Toft, Annette Hanne
APPLICANT: Shlehn, Martin
TITLE OF INVENTION: Prevention of Back-Staining
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033,537A

```

FILING DATE: 02-MAR-1998
 CLASSIFICATION: 008
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 0993/95
 FILING DATE: 08-SEP-1995
 APPLICATION NUMBER: PCT/DK96/00364
 FILING DATE: 03-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Green, Reza
 REGISTRATION NUMBER: 38,475
 REFERENCE/DOCKET NUMBER: 4492.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 551 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-033-537A-1

Query Match 7.2%; Score 189; DB 2; Length 551;
 Best Local Similarity 21.3%; Pred. No. 2.9e-08;
 Matches 108; Conservative 72; Mismatches 181; Indels 146; Gaps 24;

QY 38 VSGTLLYDANGNPFVARGI-NHGNAVYKQOATTAIEGIANGTAVRIYLSDGQWTKDD 96
 DB 11 VQGNQVGGSGQAVOLVGGSHGLOMYGN-----FNKSSLOMRRN 52
 QY 97 --IHVYRNILSLAEDNHLVAVEVDATGYDTSASLNRA---VDYWT----- 138
 DB 53 WCINFRAMRYAEIDGY-ITDPSVKKKVKEAVQASIDLGIVYIDWHIISDGNPTYKAQ 111
 QY 139 -----EMRSALIGKEDTVIINIANEMFGSMGDMAD--GY-KQALPRLNAGLNTIYAV 190
 DB 112 SKAFQEMATLGNPNVYIEIANEPNGN--VSMADVKSIAEYVTAIRADPDGVVIV 168
 QY 191 DAGMGQFQSIHDYGREYFNADPQRTMFSIHMEYAGNASSQVNTIDRYLNDLALV 250
 DB 169 GSPPTMSQ-----DIHLADNPVSHSNVYKALHF--YSGTHGFLLDRITTYANKGAALF 220
 QY 251 IEEFGHRTNGD-----VDEATIMSYSEQRGVGWLAMS-----WKGNGPEWEYL 294
 DB 221 VTEMGSDASGNGGPFYPOSKIEWIDFLNARKISVMNMSLADKVEYSAALMFGASP----- 275
 QY 295 DLSNMGAGNNTLAWGNTIYNGPYGLRETSRLSTVFTGGGSDGT---SPTTLYDFEGSMQ 351
 DB 276 --TGGWTDALQLESQKAV-----RDQIRQA--TGGSGNFTAPAAPNLNSATAGNAQ 323
 QY 352 -GWTGSSLSGGFWAVTEMSSGSHSLKADIQLSNSQHYLHYQNTSLQONSRIQATYKH 410
 DB 324 VSLTNNAVSGAT-----SYTK-----RATISG 346
 QY 411 ANMGVYGNMTARLYYKIG--HGYTWY-----SGSFVPINGSSGTTLSLDSNVQNLQV 463
 DB 347 GPTYNATGVTATSYNTGLTNGTLYYVVASNS---AGSSANSAQASATPAGSAGAST 402
 QY 464 REIGVOFO-----SASDSGQTSIYIDN 486
 DB 403 GNLVVOYKVGDTSATDNQMKPSFNKN 429

RESULT 8
 US-08-924-440-2
 ; Sequence 2, Application US/08924440
 ; Patent No. 5871550
 ; GENERAL INFORMATION:
 ; APPLICANT: Frits et al.
 ; TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
 ; NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genencor International, Inc.
 STREET: 925 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/924,440
 FILING DATE: August 27, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Christopher L. Stone
 REGISTRATION NUMBER: 35,696
 REFERENCE/DOCKET NUMBER: GC388
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 845-6504
 TELEFAX: (650) 845-7555
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 430 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-924-440-2

Query Match 6.7%; Score 174; DB 2; Length 430;
 Best Local Similarity 24.7%; Pred. No. 4e-07;
 Matches 81; Conservative 49; Mismatches 130; Indels 68; Gaps 19;

QY 26 GSTANANGSG-FYVSGTLLYDANGNPFVARGIN-HGNAVYKQOAT-TALEGIA-NTGANT 81
 DB 124 GGTGTPERYGKQVQVCIQCDHEGNPNVQRMSTHGIQWFDHCLTDSALAYDKADI 183
 QY 82 VRI---VLSDDGQW---KDIHYRNILSLAEDNHLVAVEVDATGYDTSASLNRAV 134
 DB 184 IRLSMYIOEDGYETINRGFTDRH---QLIDVATARGLYIVDWHILTFGDHYNDPRAK 240
 QY 135 DWIEMRSALIGKEDTVIINIANEMFGSMGDMAD--GY-KQALPRLNAGLNTIYAV 191
 DB 241 TFFAEIAQNHASKTN-VLEIANE-----PNGVSMASIKSYAEVLPVIRQRPDSYITVG 295
 QY 192 AAGWGQFQSIHDYGREYFNADP-QRNTMFSIHMEYAGNASSQVNTIDRYLNDLALV 249
 DB 296 TGWMSLQVVS--GSGPAEIAAPVNASNIMYAFHFY-----AASHRDYVNLNLRASL 348
 QY 250 ---VIGFGRHRTNGD-----VDEATIMSYSEQRGVGWLMSK----- 285
 DB 349 FPEVTFEFTETTYTGANDFQMAQR--YIDMAERKIGITKWNYSDDPRGSAVQPGTIC 406
 QY 286 -GNGPEWEYLDSNMGAGNNTLAWGNTI 312
 DB 407 ASGGP-----WSSSLKASGQWV 424

RESULT 9
 US-07-862-588B-7
 ; Sequence 7, Application US/07862588B
 ; Patent No. 5916796
 ; GENERAL INFORMATION:
 ; APPLICANT: Joergensen, Per Linaa
 ; APPLICANT: Sch. Iein, Martin
 ; APPLICANT: Hansen, Christian
 ; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
 ; NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 59167960 No. 5916796d1sk of No. 5916796th America, Inc.
 STREET: 405 Lexington Avenue, 62nd floor
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10017

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/862,588B
 FILING DATE: 1992/02/27

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: DR 164/90
 FILING DATE: 19-JAN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/DK91/00013
 FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:
 NAME: Zelson, Steve T. / Lambiris, Elias J.
 REGISTRATION NUMBER: 30,335 / 33,728
 REFERENCE/DOCKET NUMBER: 3425,204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 531 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-862-588B-7

Query Match 6.1%; Score 160; DB 2; Length 531;
 Best Local Similarity 20.6%; Pred. No. 8.9e-06;
 Matches 109; Conservative 73; Mismatches 214; Indels 132; Gaps 24;

QY 11 ITLSILLASLLFVSGISTANANGS-----FYVSGTITIDANGNPFVARGI-NNGHAMVK 64
 DB 12 LVLEWVLMGLLPVAGKGYAAPVPGOLKVGQGNOLVGGSGAVOLVGGSSHGLQMTG 71
 QY 65 DQATAIGIANTGANTVRIYLSGGQWTKDD-ITHVRNLISLAEDNHLVAVEVDHAT 122
 DB 72 N-----FYVKSLSQWRRDNGINVRAMTISEDGY-ITDPSYKMKV 112
 QY 123 GYDSIASLNR-----VDYWI-----EMRSALIGKEDTVIINIANEWF 161
 DB 113 KEAVQASIDIALYVLIIDWHLISDGNPTKYAKSKAFQEMATLYGNTPEVYEATSETE 172
 QY 162 SNEGAMADGYQALPRLRMNGINLTLMVDAGWGQFQSIHDYREVFNDPQRMFES 221
 DB 173 CVLGGQSS--BEVTAITSIDPDGVYVIGSPTWQ-----DILADNPVSHNWKYA 224
 QY 222 IHMYEYAGNASQVNTIDRVLNODLAVIGFGRHRTNGVDDEATIMSYSEQRGVWLA 281
 DB 225 LHF--YSGTHGQFLDRITRYANKGAALFVTEMGSDASGN----- 263
 QY 282 WSMKNGPMEYELDISNDY---AGNNLTANGTYNGPGLAETSRLSVTFGGSSDGT 338
 DB 264 -----GGP---YLPOSKKEIDFLNARKISWV---WSLADKETSALMPGASPTGA 309
 QY 339 SPTLLDFEGSMQWGTSSLSGGPMAVTEMSSKSGSHL-KADIQSSNSQHYLIQWTS 397
 DB 310 GP-----MPCNRMGKSGSSSN---PASMWRQGNPTAALPYNLSANGN-----AQYS 355
 QY 398 LQONSRIQA--TVKHAN-----WGSVAGNGMTARLYVKTG--HGYTWY-----SGSEVPFI 442
 DB 356 LFMNNAVSAGATSVYKRAATSGGYTIVDNGVATATSTNTGLNGLTYVYVVRASNS--- 411

QY 443 NGSSGTLISLISNVONLSQVREIGVQFQ-----SASDSSGQTSIYIDN 486
 DB 412 AGSSANSAQASATPASGASATGNLVQYKVGDISATDNQKPSFNIN 459

RESULT 10
 US-08-276-213-3
 ; Sequence 3, Application US/08276213
 ; Patent No. 5536655.
 GENERAL INFORMATION:
 APPLICANT: Thomas, Steven
 APPLICANT: Laymon, Robert
 APPLICANT: Himmel, Michael
 TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: National Renewable Energy Laboratory
 STREET: 1617 Cole Boulevard
 CITY: Golden
 STATE: CO
 COUNTRY: USA
 ZIP: 80401-3393

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/276,213
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: O'Connor, Edna
 REGISTRATION NUMBER: 29,252
 REFERENCE/DOCKET NUMBER: NREL IR 94-08
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303)231-1000
 TELEFAX: (303)231-1098

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 521 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 US-08-276-213-3

Query Match 5.5%; Score 143; DB 1; Length 521;
 Best Local Similarity 21.0%; Pred. No. 0.00025;
 Matches 124; Conservative 70; Mismatches 180; Indels 216; Gaps 36;

QY 32 ANSGF-FVSGTITIDANGNPFVARGIN-----HGHANYKQDQATAIEGIANATGA 79
 DB 1 AGGCIWHTSGREILIDANNVVRAGINWFGFERCNCVYVG-LMSROY-RSMIDQIKSLAY 58
 QY 80 NTVRIYLSGGQWTKDDI--ITHVRNLISLAEDNHLVAVEVDADGYSIASLNRVAVY 136
 DB 59 NTRLPYS-----DDLKPGTTPNSINFYQNN-----ODLOGLSLQYMDIVAI 103
 QY 137 -----WIEMRSALI-----GKEDTVIINIANEW 159
 DB 104 AGQIGLRIILDRHRPDCSGSALMTYSSVEATWISDQLAQRKYGNPVYVGFDLHNE 162
 QY 160 FGSWEGDAMADG-----YKQALPRLRNA--GLNHTIAY-----DAAGV----- 195
 DB 163 --PHDPACWGGGSDIDMRLAERAGNAVLAVNPMLLIIFVEGYOSYNGSDYWMGCLNQA 220
 QY 196 GQFP-----QSIHDYGREVFNDPQRMFESIHMYEYAGNASQV-RINIDRVLN 244


```

Db 221 GQPYVAVLVNPNRLVYSADYATSY---PQ--TWSDPTFP---NNMGEINWKNNGYLFN 272
QY 245 QDLALV-IGEGFH--RHTNGDVDEATIMSY---SEQRGVG---WLAMSKNGGPEMEXLD 295
Db 273 QNIAPVWVGEGFTLTQSTDTQTLKTVLYLPTAQYAGDSQFTWFMWNPDSGDTGCI- 331
QY 296 LSNMAGNNLRAKMTIYNGPYGLAETSLSTVFTGGSGDGTSP--TLIYFESMGQM 353
Db 332 LKDDW-----QTVDTV-----KDGYLAPIKSIFDPVGA--- 360
QY 354 TGSLSGPMATVETWSKSGSHLKADIDLSNSQH-----Y 389
Db 361 -----SASP-----SSQPSPSVSPSPSPSASRTPPTPTPTPTPTPTPTPTPTPT 409
QY 390 LHVIONTSLQONSRIQATVK--HANWGVNGMTARLYYKTHG---YTW-----YSG-- 437
Db 410 ASPTSPPTAASGTCTASQYVNSDW---GNGFTVAVV--TNGSVATKTKWTWTFEGNQ 465
QY 438 -----SFVPIKSSGTTLSLDSNVQNLQSVREIGVQFOSASDSSGOT 480
Db 466 TITSMNAVYQNGQSVTARNMKNYNVIOPGQNTTEGFGQASYTGSNAAPT 515

RESULT 11
US-08-604-913B-11
; Sequence 11, Application US/08604913B
; Patent No. 5712142
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.
; APPLICANT: Himmel, Michael E.
; APPLICANT: Baker, John O.
; APPLICANT: Chou, Yat-Chen
; TITLE OF INVENTION: METHOD FOR INCREASING
; TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASC II (DOS) text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/604,913B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/276,213
; FILING DATE: 15-070-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Edna M. O'Connor
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: 95-56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/384-7573
; TELEFAX: 303/384-7499
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: E1-CAT
; US-08-604-913B-11

```

```

Query Match 5.48; Score 142; DB 1; Length 358;
Best Local Similarity 23.18; Pred. No. 0.00017;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

QY 32 ANSGF-YVSGTTLVDANPNPFMNGIN-----HGHAWYKDAQNTIAEGIANGA 79
Db 1 AGGGYVWHTSGREILDANNVPRIAGINWFGFETCNVYVHG--LWSHDY--RSMIDQIKSLGY 58
QY 80 NTVRIVLSDGQWTKDDI---HTVRLISLAEDNHLVAVPEVHDATGYDSIASLRADV 136
Db 59 NTRIRPYS-----DILKRGTMNSINFQNM-----QDQGLSLQVMDKIVAX 103
QY 137 -----WIEMRSALI---GREDVIINIANEW 159
Db 104 AGQIGLRITLIDHRPDCSGQSALWYTSVSEATWISDQLALQARKGNPTVGFDLHNE- 162
QY 160 FGSWEGDAMADS-----YKQALPRLRNA--GLNHTLAV-----DAAGW----- 195
Db 163 --PHDPACWGGCGDPSIDRLAERAGNVLVSNPMLLFVEGVQSYNGSDSYWVGNGNLOGA 220
QY 196 GQFP-----QSIDYGREVENADPORWTFESIHWYETAGNASQV--RTNIDRVLN 244
Db 221 GQPYVAVLVNPNRLVYSADYATSY---PQ--TWSDPTFP---NNMGEINWKNNGYLFN 272
QY 245 QDLALV-IGEGFH--RHTNGDVDEATIMSY---SEQRGVG---WLAMSKNGG----- 288
Db 273 QNIAPVWVGEGFTLTQSTDTQTLKTVLYLPTAQYAGDSQFTWFMWNPDSGDTGCI 332
QY 289 -PEWEYLDLSDND 299
Db 333 KDDWQTVDTVKD 344

RESULT 12
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BAREKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid

```

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-038-682-2

Query Match 4.9%; Score 129; DB 1; Length 1536;
 Best Local Similarity 21.1%; Pred. No. 0.021; Indels 136; Gaps 21;
 Matches 115; Conservative 74; Mismatches 221;

1 LNNGFKKISITSLILASSILFVSGTSTANANG-----FYVSGTLLYDANGPEV 52
 DB 881 INSG-----NLTAGNIVNINAGNLVESNANFKATNFTFVNG--LFDNKGNSNI 929
 QY 53 MRGINHGAWYKDOATTALEGIANGTANTVRIYLSOGGOWTKODIHTVRNLISLAEDNH 112
 DB 930 --SIAGGARFKDIDNSKMLSTTNSSYTRITISGNITKNKNDL---NITNKGSDTEM 983
 QY 113 VAVPEVHATGYDSIAS--LNRAVDYIEMRSALIGKEDTVIINIANEWFSGWEGDAMAD 170
 DB 984 QIGDVSOKEGNETISSDKIN-----ITKQITIKAGVDGE-----NSDSD 1023
 QY 171 GYKQAIPLRLNAGLNLHMLVDAWGQFPOSIDYGREVFNADPQRTMFSIHMYEYAG 230
 DB 1024 ATNNANLTIKTELKLTODLNISGFNK-----AETIAKDSDLTIGTNSAD--GT 1072
 QY 231 NASQVRLNDRVLNODLALVIGFGRHRTNGVDVETIMSYSEORGVWLAWSKNGNPE 290
 DB 1073 NAKV-----TFNQVKDSKISADGK-----VTLHAKVETSG-----SNNMTE 1110
 QY 291 WEYLDLSNDWAG-----NLTAGMNTIVNGPYGLRETSRLSTVFTG----- 331
 DB 1111 ----DSSDNAGLITDAKNVTYNNNITSHKAVSISATSGEITITKTITINATTGNVEITA 1166
 QY 332 -----GSGDGTSPITLYDEFGSMOGWTSLSGGPAAVTEMS-----KGSHS 375
 DB 1167 QTGSIIGIESSGSVTLTATEGAL--AVSNISGNTVYTANSALTLTAGSTIKGTES 1223
 QY 376 LKADIOLS-----SNSOHLVHYONTSLQONSRIQATVKHAN---WGSVNGMTATLY 425
 DB 1224 VTTSSQSDIGTISGCTVEKATESLTQNSKIKATGEANTVSATGIGTISGNTV 1283
 QY 426 VKTGHGYTWYSGFVPINGSSG---TTLSDLSNVONLSQVREIGVQFOSASDSSGOTS 481
 DB 1284 NVTANAGDLTVNGAEINATEGAATLTSSGKLTTEASHTSAKGVNLSADDSVAGS 1343
 QY 482 IYIDNV 487
 DB 1344 INAAV 1349

RESULT 13
 US-08-302-832-2
 : Sequence 2, Application US/08302832
 : Patient No. 5603938
 : GENERAL INFORMATION:
 : APPLICANT: Barenkamp, Stephen J
 : TITLE OF INVENTION: High Molecular Weight Surface Proteins
 : TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Shoemaker and Mattare, Ltd.
 : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 : STREET: Bldg. 1
 : CITY: Arlington
 : STATE: Virginia
 : COUNTRY: U.S.A.
 : ZIP: 22202-0286
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC Compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/302,832
 FILING DATE: 16-SEP-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US PCT/US93/02166
 FILING DATE: 16-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Berleissner, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-404
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0810
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1536; amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-302-832-2

Query Match 4.9%; Score 129; DB 1; Length 1536;
 Best Local Similarity 21.1%; Pred. No. 0.021; Indels 136; Gaps 21;
 Matches 115; Conservative 74; Mismatches 221;

1 LNNGFKKISITSLILASSILFVSGTSTANANG-----FYVSGTLLYDANGPEV 52
 DB 881 INSG-----NLTAGNIVNINAGNLVESNANFKATNFTFVNG--LFDNKGNSNI 929
 QY 53 MRGINHGAWYKDOATTALEGIANGTANTVRIYLSOGGOWTKODIHTVRNLISLAEDNH 112
 DB 930 --SIAGGARFKDIDNSKMLSTTNSSYTRITISGNITKNKNDL---NITNKGSDTEM 983
 QY 113 VAVPEVHATGYDSIAS--LNRAVDYIEMRSALIGKEDTVIINIANEWFSGWEGDAMAD 170
 DB 984 QIGDVSOKEGNETISSDKIN-----ITKQITIKAGVDGE-----NSDSD 1023
 QY 171 GYKQAIPLRLNAGLNLHMLVDAWGQFPOSIDYGREVFNADPQRTMFSIHMYEYAG 230
 DB 1024 ATNNANLTIKTELKLTODLNISGFNK-----AETIAKDSDLTIGTNSAD--GT 1072
 QY 231 NASQVRLNDRVLNODLALVIGFGRHRTNGVDVETIMSYSEORGVWLAWSKNGNPE 290
 DB 1073 NAKV-----TFNQVKDSKISADGK-----VTLHAKVETSG-----SNNMTE 1110
 QY 291 WEYLDLSNDWAG-----NLTAGMNTIVNGPYGLRETSRLSTVFTG----- 331
 DB 1111 ----DSSDNAGLITDAKNVTYNNNITSHKAVSISATSGEITITKTITINATTGNVEITA 1166
 QY 332 -----GSGDGTSPITLYDEFGSMOGWTSLSGGPAAVTEMS-----KGSHS 375
 DB 1167 QTGSIIGIESSGSVTLTATEGAL--AVSNISGNTVYTANSALTLTAGSTIKGTES 1223
 QY 376 LKADIOLS-----SNSOHLVHYONTSLQONSRIQATVKHAN---WGSVNGMTATLY 425
 DB 1224 VTTSSQSDIGTISGCTVEKATESLTQNSKIKATGEANTVSATGIGTISGNTV 1283
 QY 426 VKTGHGYTWYSGFVPINGSSG---TTLSDLSNVONLSQVREIGVQFOSASDSSGOTS 481
 DB 1284 NVTANAGDLTVNGAEINATEGAATLTSSGKLTTEASHTSAKGVNLSADDSVAGS 1343
 QY 482 IYIDNV 487
 DB 1344 INAAV 1349

RESULT 14

US-08-530-198-2
Sequence 2, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME IIT, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530.198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-2

Query Match 4.9%, Score 129, DB 2, Length 1536;
Best Local Similarity 21.1%, Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;
DB 1 LUNGFKKIFSLILLASSILFVSGSTANANSG-----FYVSGTLLYDANGNPFV 52
DB 881 INSG-----NLTAGNIVNINAGNLVESNANFKATINTFTVNG--LFDNKGNSNI 929
QY 53 MRGINHGWKQDQATTAIEGIANTGANTVRIVLSDGGQWTRDIIHTVNLISLAEDNL 112
DB 930 --SIKAGARFDIDNSKMLSTTNSSSYRTIISGNITNKNGDL---NINNESDTEM 983
QY 113 VAVPEYHATGDSIAS--LNPADVWIMKRSALIGKEDTVIINTANEFVGSWEGDAMD 170
DB 984 QIGGVSOKEGRLTISSDKIN-----ITKQITKAGVDGE-----NSDSD 1023
QY 171 GYKQAIPLRNALNHTLPLVDAAGQFPQSIHDYGREVFNDPQNTMESIHMYEYAG 230
DB 1024 ATNNANLTKTEKLTDODLINSFKN-----AETLADGSDLTIGNTNSAD--GT 1072
QY 231 NASQVRTNIDRYLNDOLALVIGFGRHRTNGDVDEATINSYSEQRGVGLAWSKNGPE 290
DB 1073 NAKVY-----TFNQVKDSKISADGK-----VTLHSKVTESG-----SNNTE 1110
QY 291 WEYLDLSDMAG-----KNLTAMGNTIVNGPYGLKETSRLSVTFG----- 331
DB 1111 ---DSDNNALTLTDARKVYNNNITSKAVISATSGEITTKGTITNATGVEIRA 1166
QY 332 -----GGSDGTSPTLLYDFEGSGMGWGTGSSLGSPNATVEMSS-----KGSMS 375

DB 1167 QNGSILGIESSSGSVTLTATEGAL---AVSNISGNTVTVANSGALLTLGASTIKGES 1223
QY 376 LKADIQLS-----SNSQHYLVIONTSLOQNSRIQATVKAN---NGSVNGKATRLY 425
DB 1224 VTTSSQSGDIGGTIGSGYEVATKESLTQNSKIKATGAEVNTSANGTIGTISGTV 1283
QY 426 VATGHTYMYSGSFYPINSSG-----TTLSDLSVQVONLSQVRELVGFQSSDSSGTS 481
DB 1284 NYTANAGDLTVNGAEINATGGAATLTYSGRKLTREASHITSAGQVNLASQDSVAGS 1343
QY 482 IYIDNV 487.
DB 1344 INANV 1349.

RESULT 15
US-08-469-880-2
Sequence 2, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: OF No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469.880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bernstreser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:Vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2
Query Match 4.9%, Score 129, DB 2, Length 1536;
Best Local Similarity 21.1%, Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;
DB 1 LUNGFKKIFSLILLASSILFVSGSTANANSG-----FYVSGTLLYDANGNPFV 52

Db 881 INSG-----NLTAGNTVNIAGNLTVESNANFEAITNFTENVGG--LEFDNKGNSNI 929
QY 53 MRGINHGHWYRDQATTAIEGIANTGANTVRIVSDGQWTDIDHVRNIIISLAEDNHL 112
Db 930 --SIKGGARFEDINDSKNLSTTNSSSTYRIISGNITNKGDL---NITNKGSDTEW 983
QY 113 VAVPEVHATGDSIAS--LNRADVWIEMRBALIGKEDTVIINANEMFGSWEGDAMAD 170
Db 984 QIGGDVOREGNLTSSDKIN-----ITKQITINAGVDGE-----NSDSD 1023
QY 171 GYKQAIPLRLNAGLNHTLWVDAAGWGPQSIHDYGREYFNADPQANTWFSIHMEYAGG 230
Db 1024 AITNANLTIKTEKLTODLNISGFNK-----AETAKGSDLTIGNTNSAD--GT 1072
QY 231 NASQVFNIDRYLNODLALVIGFGRHRTNGVDENTISYSEQRGVGMWAMSKNGNPE 290
Db 1073 NAKKV-----TFNQYKDSKISADGK-----VTLHSEVETSG-----SNNITE 1110
QY 291 WEYLDLSNDWAG-----NNLTAWGNTIVNGPYGLRETSRLSTVETG----- 331
Db 1111 ----DSSDNAGLTIDAKNVTYNNNITSHKAVISATSGEITTKGTINATGNVEITA 1166
QY 332 -----GSGDGTSPPTLYDFEGSMQWGTGSSISGCPNAVTEWS-----KGSHS 375
Db 1167 QTSILGIESSSGSVTLTATEGAL--AVSNISGNTVTVTANSALTTLAGSTIKGTES 1223
QY 376 LKADIDQLS-----SNSQHLVIONTSLOONRRIQATVKHAN---WGSVGNMGTARLY 425
Db 1224 VTTSSQSGDIGTISGTVVEVKATESELTTOSNKRKATIGEANVTSATGTIGTIGNTV 1283
QY 426 VKTGHGTYWYSGFVPINGSSG---TTLSLDLSNVQNLQVREIGVQFQASDSSGQTS 481
Db 1284 NVTANAGDLVVGAGAEINATEGAATLTSSGKLTTEASHITSAKGVNLSAQDGSVAGS 1343
QY 482 IYIDNV 487
Db 1344 INANV 1349

Search completed: December 19, 2000, 17:02:41
Job time: 1273 sec

QY 421 TATLYKTHGHTWYSGSPVINGSGTTISLDSNVQNLISQVREIGVQFQASDSSGQT 480
 DB 447 YATLYKTHGHTWYSGSPVINGSGTTISLDSNVQNLISQVREIGVQFQASDSSGQT 506
 QY 481 SITIDNIVE 490
 DB 507 AITVDVSTIQ 516

RESULT 2
 T37232
 secreted beta-mannosidase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 A:Accession: T37232
 R:Oliver, K.; Harris, D.; Parrhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: 221615
 A:Accession: T37232
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-384 <OLI>
 A:Cross-references: EMBL:AL031514; PIDN:CAA20610.1; GSPDB:GN00070; SCODEB:SC2H4.16
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: manA2; SCODEB:SC2H4.16

Query Match 29.2%; Score 761.5; DB 2; Length 384;
 Best Local Similarity 44.2%; Pred. No. 19.8e-45;
 Matches 161; Conservative 59; Mismatches 121; Indels 23; Gaps 8;

QY 9 FSTTSLASSILFVSGTSTANANS---GFYVSGTIVDANGNPVINGINGHAMTKD 65
 DB 16 FAIVLGLFRA-----LAGPSAGRAEAAGGIHVSNGRVLEGNSVFMKRVNAIYTPDR 70
 QY 66 QATTALEGANTGANTVRIYSDGQWTKDITVRNLISLAEDNHLVAVPEVHDATGY- 124
 DB 71 R-TGSTADIADAKANTVRYVLSGSGRMTKSASEVSAIIQCCANKNYICYLEVHDTTGYG 129
 QY 125 --DSIASLADVWIEKRSALIGKEDVTIINANEFWSGEGDANADGYKQAIPLRNA 192
 DB 130 EDAANAATSLDQADADYVSVKSALEGQEDYVVYVNGNEPFGNTNTAMTDATKSAIGKLGA 189
 QY 183 GLNHTLVADAAGGQ--FPOSIHGYREVFNADQRMNMFISHYEYKAGNASOVRTNIDR 241
 DB 190 GLDHALVADPANNQGDQSGTMRNSAASVFAASDDPRTVFSVHYG-VYDTAAEVDYLA 248
 QY 242 VLNQDLALVIGFGRHRTNGVDVDEATIMSSEQGVGLAMSKNGKPEWEYLDLSNDMA 301
 DB 249 FVSGSLPIVGEFGDSDNDPDEDAIMATQAQSLGVLGMSNGSGGVETIDMNVGFD 308
 QY 302 GNNLTAMGNTIVNGPYGLRTSRLSYF-TGGGSDGTSPT-----TLVDFEGSMOG 352
 DB 309 PNLSTGNGRIFFYGSNGIATSRATATVYGGGGSTGTATNGPYCYVNGASPDGDMG 368
 QY 353 WTGS 356
 DB 369 WENS 372

RESULT 3
 S30386
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) precursor - Streptomyces lividans
 N:Alternate names: beta-mannanase
 C:Species: Streptomyces lividans
 C:Date: 02-Dec-1993 #sequence_revision 13-Mar-1997 #text_change 22-Oct-1999
 A:Accession: S30386; S27699
 R:Archand, N.; Kluepfel, D.; Paradis, F.W.; Morosoli, R.; Shareck, F.
 Biochem. J. 290, 857-863, 1993
 A:Title: beta-Mannanase of Streptomyces lividans 66: cloning and DNA sequence of the man
 A:Reference number: S30386; MUID:93207541
 A:Accession: S30386

A:Molecule type: DNA
 A:Residues: 1-363 <ARC>
 A:Cross-references: EMBL:M92297; NID:g153193; PIDN:AAA26710.1; PID:g404076
 A:Experimental source: strain 1326
 C:Genetics:
 A:Gene: manA
 C:Function:
 A:Description: catalyzes degradation of beta-mannans into short-chain oligo-mannoside
 A:Pathway: mannose metabolism
 A>Note: Industrially used for bleaching of wood pulps
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F.1-36/Domains: signal sequence #status predicted <SIG>
 F.37-363/Product: mannan endo-1,4-beta-mannosidase #status predicted <MAT>

Query Match 28.1%; Score 731.5; DB 2; Length 363;
 Best Local Similarity 46.1%; Pred. No. 1e-42;
 Matches 155; Conservative 55; Mismatches 111; Indels 15; Gaps 8;

QY 13 LSTLLASSILFVSGTSTANANS---GFYVSGTIVDANGNPVINGINGHAMTKDQAT 69
 DB 19 LGILFRA-----LAGPSAGRAEAAGGIHVSNGRVLEGNSAVFMKRVNAIYTPDR-TG 72
 QY 70 ALEGANTGANTVRIYSDGQWTKDITVRNLISLAEDNHLVAVPEVHDATGY--DSI 127
 DB 73 STADIAKANTVRYVLSGSGRMTKSASEVSAIIQCCANKNYICYLEVHDTTGYGKGA 132
 QY 128 ASINRAVDYIEMKRSAL-IGKEDVTIINANEFWSGEGDANADGYKQAIPLRNAGLNH 186
 DB 133 TSDIQAGDVTIVGKSAAMRAQEDYVVYVNGNEPFGNTNTAMTDATKSAIGLRGKGLH 192
 QY 187 TLMVDAAGGQ--FPOSIHGYREVFNADQRMNMFISHYEYKAGNASOVRTNIDRLNQ 245
 DB 193 ALKAVDAPNNQGDQSGTMRNSAASVFAASDDPRTVFSVHYG-VYDTAAEVDYLAFCVN 251
 QY 246 DLALVIGFGRHRTNGVDVDEATIMSSEQGVGLAMSKNGKPEWEYLDLSNDMAGNNL 305
 DB 252 GLPIVGEFGDSDNDPDEDAIMATQAQSLGVLGMSNGSGGVETIDMNVGFDPNL 311
 QY 306 TANGNTIVNGPYGLRTSRLSYF-TGGGSDGTSPT 340
 DB 312 TSWGNRLYGSNGIATSRATATVYGGGGSTGTATP 347

RESULT 4
 A48954
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocoellum saccharolyticum
 N:Alternate names: beta-mannanase
 C:Species: Caldocoellum saccharolyticum
 C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
 A:Accession: A48954; B43745
 R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
 Appl. Environ. Microbiol. 58, 3864-3867, 1992
 A:Title: The beta-mannanase from "Caldocoellum saccharolyticum" is part of a man
 A:Reference number: A48954; MUID:93119139
 A:Accession: A48954
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-1331 <GIB>
 A:Cross-references: GB:I01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
 A>Note: sequence extracted from NCBI backbone (NCBI:121576; NCBI:121577)
 R:Luehl, E.; Bhana Jaemai, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
 Appl. Environ. Microbiol. 57, 694-700, 1991
 A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene
 A:Reference number: A43745; MUID:91247819
 A:Accession: B43745
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-337, 'PPOHORO' <LUE>
 A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
 A>Note: the authors translated the codon CAC for residue 262 as Glu
 A>Note: this sequence has been revised in reference A48954
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

[illegible]

Q:Function: cleaves the beta-1,4-mannosidic linkages in various beta-mannan
 C:Description: cleaves the beta-1,4-mannosidic linkages in various beta-mannan
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 16.48; Score 427; DB 2; Length 513;
 Best Local Similarity 26.98; Pred. No. 9,66-22;
 Matches 149; Conservative 76; Mismatches 215; Indels 114; Gaps 23;

QY 5 FKKIFSTLCLLASSILFVSGSTANANSGPYVSGTLYDANGNPFVW-----53
 DB 4 YKKVAFVAAIFEEFVLPTLTISSMSSEAN-----GAALSNPNANQTKKNYSWLANLPN 55
 QY 54 ---RQINNGH-AMYDQATTAIEGIANGTANTVRIYS-DGGQW-----TKDDIHTVAN- 102
 DB 56 KSNKRVVSGHGEYSDSTLWATKOCARELTGKMPGLSCDRIKKNMOTRLYLVAQSYGCNQ 115
 QY 103 -LISLAEDNHLAAV-----PEVDATGYDSIASINRAVDY-----WIEMRSALI 145
 DB 116 ELINFWNNOGGLVTVISVHPNPGFHSGENYKTLPTSQONLNNHTTEGRKKKMDLAKA 175
 QY 146 GKED-----TVIINIANEFWSWEGDAW-ADGYKQAIPLRLNAGLN-----H 196
 DB 176 DGLDELQNNGVVLEFRPLHEMNGEW---FWMGAEQYNOFDQIRANAYIASMRDMYQYFTH 232
 QY 187 TLMDVAGKGPQPSIHGYGREYVADPQKNTMSEIHMYEYRAGASQVPRNIDPVLNQD 246
 DB 233 ERKLNKLWVYSPDYRPH---VTSYYPGANYVDLVALDSY-HPPHSLTDQYNMIALD 288
 QY 247 LALVIGEGRHRT-NGDVDEATIMSSEQ---RGYGLWMS--RK--GNGPWEYLDLSN 298
 DB 288 KRFAPAEIGPFSMAGSDYSNYIQAIRKPKRYTFYFLAMDKMSPHNRCAM---DLFN 345
 QY 299 DWAGNNLTAMGNTIVNGPYGLRETSRLSTVFEGGSDGTS--PTLYDFEGSMQGTGS 356
 DB 346 D-----SMVYVRGELIDYGQSNPAPVLYDFENNTLSWSGC 379
 QY 357 SLIS-GGPAVTEWSSKSGSHSLKADIQLSSNSOHIYHVIQNTSLQONSKIQATVKANNGS 415
 DB 380 EPTDGGPMTSNEMSNAGTOSLKAADVLLGNNSYH-LQKTVNRNLSFKMLKIVSHSSWGN 438
 QY 416 VNGWTAIRLYVTGIGYMYSGSFVPINGSSGTTLSLDSNNQNSOYREIGVQPOSASD 475
 DB 439 VSGSTARFVFTGSAMRNWAGFCQFAGKRTTALSLDILTKVNLHDVREIGVEYKAPAN 498
 QY 476 SSGQSTIYDNIYV 489
 DB 499 SNGKTAIYDHYTV 512

RESULT 6 ;
 JCS487
 cellulase (EC 3.2.1.4) precursor - Clostridium cellulovorans
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: Clostridium cellulovorans
 C:Date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 15-Oct-1999
 C:Accession: JCS487; .PC4333
 R:Shewetta, S.A.; Ichihashi, A.; Park, J.S.; Liu, C.; Malburg, L.M.; Doi, R.H.
 Gene 182, 163-167, 1996
 A:Title: Characterization of enfF, a gene for a non-cellulosomal Clostridium cellulovorans
 A:Reference number: JCS487; .MUID:97136706
 A:Accession: JCS487
 A:Molecule type: DNA
 A:Residues: 1-557 <SHE1>
 A:Cross-references: GP:U37056; NID:g1778708; PIDN:AMB40891.1; PID:g1778709
 A:Experimental source: strain ATCC 35296
 A:Accession: PC4333
 A:Molecule type: Protein
 A:Residues: 30-37 <SHE2>
 C:Comment: This enzyme plays a secondary role in cellulose degradation.
 C:Genetics:
 A:Gene: enfF
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-29/Domains: signal sequence #status Predicted <SIG>
F:30-557/Product: cellulase #status Predicted <MAT>

Query Match 9.2%; Score 239.5; DB 2; Length 557;
Best Local Similarity 23.1%; Pred. No. 6.7e-09;
Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

```

QY 2 NNGFKRIFSI-----TLSSLLASSILFVSGTS-----TANAN-----SGFTV 38
DB 3 NNVKRIILSIYVAGAGALMLVPIVINYAAETYSNLGNNAVKPKPSYGLKQLINKNGI-- 60
QY 39 SGTLLIYDANGNPFVNGIN-HGHAWTKDQATTAIEGIANTGA-----NTVATV- 86
DB 61 --KTLCDKGNPIQLKGMSTHGLQW-----PGVANNNAFALSNOWNSNVIRLAMY 110
QY 87 -SDGQWMTDDI-HYVRNLISLAEDNHLVAVPEVHDATGYDSTASLNRAVDYWIEMRSAL 144
DB 111 VAEGGYATNPVKQYVYINQINIAANDMVIVDMHMANPDNNAASYSGAQSFNDISTYL 170
QY 145 ICKEDVIINIANEMFGSWEG--DA--WADGYKQALP--HLRNAGLNHTLMVDAAGMG 196
DB 171 YNNKRIITELCNERPENGEGVTNDATGAQYKSYATPIVOLLKNGENELITVGNPFWS 230
QY 197 QPQSIHDYGREVFADPQRTMFSIMHYEYAGN-ASQVETNIDR-----VLNODIA 248
DB 231 QRPDLAAD-----NPINDSNTMYVHF--YSGTNPISIVDNPBDAAGNVRYALNHGA 282
QY 249 LVIGEFGHHTNGVDENATIMSYSEGRGVGLAMSKMGKGEWEYLDLSND----- 299
DB 283 VYATWEGTSLATG-----TGPLYL-AKADAWLDF-LNGNNISMCFEISLNDEKAALNS 335
QY 300 -----WAGNLPLZAMG-----NTIVNGPYGLRSTSLSTVETGG 332
DB 336 LLSLDPGSKLMADEMLTSSGQYVARIRIGAYTATPVDTYQTPAPKPFSSGFDFENDG 395
QY 333 GSDG-----GTSPTLYDEFGSMQGTGSLSGPMAVTEMSSKSHSLK-----ADIOLS 383
DB 386 TTQEGGVNPDSPITAIINVENANNAALKISNL-----NSKGSNDLSEGNFMANVRIS 445
QY 384 SNS-QHYLVHVIQNTSLQONSRIQATYKHAHMGSVGNGMTARLYVTGHHY-----TW 434
DB 446 ADIWQOSINITYGDTKILMDVYIAPTPV-----NVSIAALPQSTHGMNPTAIRW 496
QY 435 YSGSEVPINGSSGTTSLDLSNVQNLQYREIGVQFOSASDS-----SGQSTIYI 484
DB 497 TNNFYAQQDGYKATLTITSTNPNFNITIA-----TDAADSVYNNMILFVGSNDNISL 550
QY 485 DNV 487
DB 551 DNI 553

```

RESULT 7

A27631
cellulase (EC 3.2.1.4) precursor - Clostridium acetobutylicum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium acetobutylicum
C:Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.
A:Title: Structure of an endo-beta-1,4-glucanase gene from Clostridium acetobutylicum P2
A:Reference number: A27631; MUID:88268074
A:Accession: A27631
A:Molecule type: DNA
A:Residues: 1-448 <ZAP>
A:Cross-references: EMBL:M31311; NID:q144789; PIDN:AAA23230.1; PID:q144790
A>Note: The authors translated the codon GAG for residue 116 as Gly, GAA for residue 263
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 9.0%; Score 234; DB 2; Length 448;
Best Local Similarity 22.6%; Pred. No. 1.2e-08;
Matches 115; Conservative 75; Mismatches 185; Indels 134; Gaps 24;

```

QY 3 NNGFKRIFSI-----TLSSLLASSILFVSGTS-----TANAN-----SGFTV 38
DB 9 NFKKTFSEFLAVN-----MLFTVLSTNTYKAAEATTSGGQLKVVGSQDCSNGKPIQLK 65
QY 55 GI-NHGHAWY-----KDQATTAIEGIANTGANTVATV-SDGQWMT--KDDI 97
DB 66 GMSHGLQWYVYVYDGMKFLRDK-----WGYNVIRAMYTEGEGYISNPSSQK 115
QY 98 HVRNLISLAEDNHLVAVPEVHDATGYDSTASLNRAVDYWIEMRSALIGKEDTYINIAN 157
DB 116 EKIKRIYODADILNMYIIVHILSDNNPNTYKQKAKSFPEMAEE-YGKTSNYIYEICN 174
QY 158 EWFSGWEDDANADYK-----QAIPLRNAGLNHTLMVDAAGMGPQSIHDYGREVFAD 213
DB 175 EPNG-----GTNNANDIKPYANTYIPAIRAIDPNNTIIVGTSWSQDVDAAD-----N 223
QY 214 POR--NTFESTHMEYAGNASQVETNIDRYLNDLALVIGFGRHRTNGVDENATIMSY 271
DB 224 PLRYSNIMYTFHF--YAGTHQSLRDKINYAMSKGIAIFTEWGTSDASGN----- 272
QY 272 SBQRGVGLAMSKMGKGEWEYLDLSNDW-----AGNNTLANGNTIVNGPYGLRSTSLST 327
DB 273 -----GGP-----YLDESQKWDVPMASKNI-SWTN-----WALCDKSEASA 307
QY 328 VFTGGSDGTSPTLYDEFGSMQGTGSSLS-----GPMVAVTEMSSKSHSLK 377
DB 308 ALKSGSSTTG-----GWDSDLTITGLFVKKSIGG-----SNTTSQISA 346
QY 378 ADIOLSSNQHYLVHVIQNTSLQONSRIQATYKHAHMGSVGNGMTARLYVK--TGHHYTW 434
DB 347 PTFSLQSGTYSAQTVLTSLSDNSDIHYTDTGTPSSPYVTSPISTIKTYKAFET 406
QY 435 YSGSEVPINGSSGTTSLDLSNVQNLQY 463
DB 407 KTG-----MTDSNITSVAYTITSNTPVKOV 431

```

RESULT 8

A25156
cellulase (EC 3.2.1.4) 1 - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
R:Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic Bacillus sp. s
A:Reference number: A91825; MUID:87056924
A:Accession: A25156
A:Molecule type: DNA
A:Residues: 1-488 <FUK>
A:Cross-references: GB:M4781; GB:X53449; NID:q142659; PIDN:AAA22301.1; PID:q142660
A:Experimental source: Strain N-4, plasmid pNK1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 8.7%; Score 226.5; DB 2; Length 488;
Best Local Similarity 23.1%; Pred. No. 4.3e-08;
Matches 98; Conservative 67; Mismatches 157; Indels 103; Gaps 22;

```

QY 6 KKSISLSSLLASSILFVSGTSAN-----ANSGFYSGTLLIYDANGNPFVWKG-I-NH 58
DB 2 KKLITLITVFTLA--LLEVGSTSNANGSVVEONGQLSIQNGOLVNHGDPVOLKGMSSH 59

```


RESULT 11

cellulase (EC 3.2.1.4) precursor - *Erythrina carotovora* (SCC 3193)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Erythrina carotovora*

A:Accession: SCC 3193

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

Mol. Gen. Genet. 247, 17-26, 1995

A:Title: Structure and regulation of the *Erythrina carotovora* subspecies *carotovora* SCC3193

A:Reference number: S54744; M01D:95231512

A:Residues: 1-504 <MAE>

A:Gene: celV1

C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-39/Domain: signal sequence #status predicted <SIG>

F:33-504/Product: cellulase #status predicted <MAT>

Query Match

Best Local Similarity 7.9%; Score 207; DB 2; Length 504;
Matches 110; Conservative 63; Mismatches 191; Indels 130; Gaps 22;

```

QY 6 KRFTSTLLSLASSLFLFSGSTANANG-FYVSGTLLYDANGNPFVNGI-NHGMAY 63
DB 10 RLTGCVVTMLGMSLFSALSTPEYTHGQSLTEGRLDEGKRYQLGSSISLQWY 69
QY 64 KQATTAIEGANTGANTVATVLDGQWTKD-IHTVKNLISLAEEDNHLVAPEV--- 118
DB 70 GD-----YVKNDSKMLRDMGIVNFRVATAE-NGYIANPILANK 110
QY 119 -----HDATGYDSIASLNAVDTWIMRSALIGKEDTVIINANEM 159
DB 111 VKEAVAAAGLGYIIIDWHTLLSDNDPTTKAQAIFFAEM-AGLGNSPNVIETLANEP 169
QY 160 FGS--WEGD--AWADGKQAIPLRNAGLHTLVDAAGQFPOSIDHYGREVFADPQ 215
DB 170 NGSVTNGQIRPYA---LETDITRSKDPNLLIVSGSTMS---QDIHD---AADNOLD 220
QY 216 RNTMFSIHMTYAGNAGSOVRTNIDRYLNDLALVIGEGHRTND-----VDEATIMS 270
DB 221 PNTLYALHF--YAGTHGOFLRDRIDYAGSRGAIIFVSEMGTSDASGNGGFLPESETWID 278
QY 271 YSRQRCVGLAMSKWKGPEWEL---DLSNDMAGNNLTAMNTLYNGPYGLRETSRLS 326
DB 279 FLNNRGISVWMSLSDKSETSAALVAGASKSGGNTDONLSTSGFY-----REQIR-- 329
QY 327 TVEFGSGDGTSPPTLYDEFGSMQGTGSLSGPWAIVEMSSKSHSLKADLOLSNS 386
DB 330 ---AGAGLSGDDPTPTMTEPTNPGNGTIG-----DIYL----- 389
QY 387 OHYLVHVIQNTSLQONSTQIAIVKANKSVG---NGMTALYKYGITGTYGSGSEPPIN 443
DB 360 -QYRNVNDNPS--DDAIRAFANIKNTGSPRIKLSLDQVYTFHD-----D 401
QY 444 GSGGTSLDLSNV 457
DB 402 GKPGANLFVDNAVY 415

```

RESULT 12

cellulase (EC 3.2.1.4) precursor - *Pseudomonas fluorescens*

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Pseudomonas fluorescens*

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S56132; S56127; S54798

R:Hall, J.; Black, G.W.; Ferreira, L.M.A.; Millward-Sadler, S.J.; All, B.R.S.;

Biochem. J. 309, 749-756, 1995

A:Title: The non-catalytic cellulose-binding domain of a novel cellulase from

A:Reference number: S56127; M01D:95369948

A:Accession: S56132

A:Molecule type: DNA

A:Residues: 1-570 <HML>

A:Cross-references: EMBL:X86798; NID:9806573; PIDN:CAA60493.1; PID:9806574

A:Accession: S56127

A:Molecule type: protein

C:Genetics: A:Gene: celE

C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: bacterial cellulose-binding domain homology; glycosidase GWM domain h

C:Superfamily: bacterial cellulose-binding domain homology; glycosidase GWM domain h

F:1-39/Domain: signal sequence #status predicted <SIG>

F:40-570/Product: cellulase #status experimental <MAT>

F:380-417/Domain: glycosidase GWM domain homology <GWM>

F:472-569/Domain: bacterial cellulose-binding domain homology <BCB>

Query Match

Best Local Similarity 7.8%; Score 203.5; DB 2; Length 570;
Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

```

QY 14 SLLASSLFLFSGSTANANGFYVSGTLLYDANGNPFVNGI-N-----HGHWTKD 65
DB 22 AILLGSGGLGVAQADVAFLSVQGNKIL-ANGQAFSGMSLFMSNTMEGGEKYYNA 80
QY 66 QATTAIEGANTGANTVATVLDGQWTKD-IHTVKNLISLAEEDNHLVAPEV 118
DB 81 QVSWEL--SDMNAVLPAAGVDEGGYITDPANKQVTOYVD-AAIANDMYIIMHS 137
QY 119 HDATGYDSIASLNAVDTWIMRSALIGKEDTVIINANEMFG-SWEGDAAWADGKQA-I 176
DB 138 HNAHQYQS---QALAFQEQE-ARKYGANNHVIYEINEPLQVMSWT--IKPYAQAVI 189
QY 177 PLRNAGLHTLVDAAGQFPOSIDHYGREVFADPQ---QRTMFSIHMTYAGNAGS 233
DB 190 AAIRAIDPNDLITVGTPTMSQ-----DVDVANDITGYONTAYTLHF--YAGTHGO 239
QY 234 QYRTNIDRYLNDLALVIGEGHRTNDVDYDQATINSYSEQGVGLAMSKWKGPEWEL 293
DB 240 YLRDRAQRLANGIALVTEMGSVANND-----GAVANSE----- 275
QY 294 LPLSDMAGNNLTAMNTLYNGPYGLRETSRLSTVFTGGSDGTSPTLYDEFGSMQGW 353
DB 276 ---TNAVVSFKMT---NHISNANMALNDKVEGASALVPGASANG-----GW 315
QY 354 TGSSTL-SGGPFA--VTEM---SSKSHSLKADLOLSNSQHYLVHVIQNTSLQONSTQ 405
DB 316 VNSQLTASGALAKSLISGPPSYNTSSSAVSSQTVQVSSSO--APVSSSSSTKSSVYS 373
QY 406 ATYV--KHANWGSVGNMGTARLYVKTGHGYTW-----YSGSFYPINGSSTTSL 452
DB 374 SAVSGQCMW---YGTLYPLCSTTTNGGWMENNSCIARATCSQGPAPWGLVGSSTSS- 428
QY 453 DLSNVNLSQVREIGVQFQSDSSSGQTSI 482
DB 429 -----QASSSVYRSSSSSLVSSRSSSSSSSV 453

```

RESULT 13

cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain IF03034)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Bacillus subtilis*

C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999

R:Nakamura, A.; Uozumi, T.; Beppu, T.
 Eur. J. Biochem. 164, 317-320, 1987.
 A:Title: Nucleotide sequence of a cellulase gene of *Bacillus subtilis*.
 A:Reference number: A27198; MID:87190397
 A:Accession: A27198
 A:Molecule type: DNA
 A:Residues: 1499 <NAX>
 A:Cross-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671
 A:Experimental source: strain IF03034
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 C:Keywords: cellulose degradation
 C:Pathway: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 8.5e-06;

Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

```

QY 6 KRIFFSTLSLLASSILF-----VSGTSTANNSG-FYVSGTTLTDANGNPFYMRG 55
DB 2 KRISIFITLLAVLTMGGLSPASAGCTKTPAAKNGOLSIKGTQVNRDKAVALKG 61
QY 56 I-NHGAWKYDQATTAEGTANTGANTVRLVSDGGQWTKDD-----IHT----- 99
DB 62 ISSHGLQWTD-----FVNKDSLKLWLDMDGIVFRAMATYADGGYI 103
QY 100 -----VANLSLAEDNHLVAPVEYHDATGDSIASLNRADVNIEMRSALIGKEDY 151
DB 104 DNPVKNKKEAVEAKELGIYIIDWHLNDGNPNQNEKAEFFKEM-SSLYGNTPNV 162
QY 152 IINIANEMFG--SWEQD--AMADGYKQALPRLNAGLNHTLMDAGMGQFPOSIDHYGR 207
DB 163 IYELANEPNDGVNKKRDKIKYAE---EVISYIRKNDPDNIIV---GTGWSQDVND--- 213
QY 208 EVENADPO---RMTMFSIHMYEYAGNAGSQRINIDRVLMQDLALVIGEGHRTNGD-- 262
DB 214 ---AADDQLDANVMYALHF--YAGTHGQSLRDKANALSKAPIFTEWGTSDASGNGG 268
QY 263 --VDEATIMSSEQGVGLWMSKNGNGPEM--EYLDLSN--DMAGNNLTANGNTIVNGPYG 318
DB 269 VFIDQSR-----EMVNTLDSKNISWVWNL----- 293
QY 319 LRETSRLSTFTGGGSDGTSPTTLTDFEGSKOGWTSSLSGGPAAVTESSKGSLSKA 378
DB 294 -----SDKQSSSSALK-----PGASKTGG-WPLTDLTASGTF-VRE 327
QY 379 DI--QLSSNQHLYLVQNTSLQONSRIQATVYHANWGSVNGMTALVYKTHGTYR--- 433
DB 328 NILGNKDSIERPPTPAQNDPADNG-ITSVQYKAGDGVNSNQTLPQLHINKNNGMATVDL 386
QY 434 -----WYSGSEVPINGSSGTTSLDLSNV 457
DB 387 KDVTARYWYNAK-----NKGQNFDCDYAQI 411

```

RESULT 14

A26874
 cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain DLG)
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 08-Mar-1989 #sequence, revision 08-Mar-1989 #text, change 15-Oct-1999
 C:Accession: A26874; B26874
 R:Robson, L.M.; Chambliss, G.R.
 J. Bacteriol. 169, 2017-2025, 1987
 A:Title: Endo-beta-1,4-glucanase gene of *Bacillus subtilis* DLG.
 A:Reference number: A26874; MID:87194581
 A:Accession: A26874
 A:Molecule type: DNA
 A:Residues: 1-508 <ROB2>
 A:Cross-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008
 A:Experimental source: strain DLG

A:Accession: B26874
 A:Molecule type: protein
 A:Residues: 39-53 <ROB2>
 A:Experimental source: strain DLG
 A:Note: the authors believe Met-1 and Met-2 may be alternate initiators
 C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase;
 F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 7.2%; Score 188.5; DB 2; Length 508;
 Best Local Similarity 21.1%; Pred. No. 1.8e-05;

Matches 78; Conservative 75; Mismatches 133; Indels 83; Gaps 17

```

QY 6 KRIFFSTLSLLASSILF-----VSGTSTANNSG-FYVSGTTLTDANGNPFYMRG 55
DB 11 KRISIFITLLAVLTMGGLSPASAGCTKTPAAKNGOLSIKGTQVNRDKAVALKG 70
QY 56 I-NHGAWKYDQATTAEGTANTGANTVRLVSDGGQWTKDD-----IHT----- 99
DB 71 ISSHGLQWTD-----FVNKDSLKLWLDMDGIVFRAMATYADGGYI 112
QY 100 -----VANLSLAEDNHLVAPVEYHDATGDSIASLNRADVNIEMRSALIGKEDY 151
DB 113 DNPVKNKKEAVEAKELGIYIIDWHLNDGNPNQNEKAEFFKEM-SSLYGNTPNV 171
QY 152 IINIANEMFG--SWEQD--AMADGYKQALPRLNAGLNHTLMDAGMGQFPOSIDHYGR 207
DB 172 IYELANEPNDGVNKKRDKIKYAE---EVISYIRKNDPDNIIV---GTGWSQDVND--- 222
QY 208 EVENADPO---RMTMFSIHMYEYAGNAGSQRINIDRVLMQDLALVIGEGHRTNGD-- 262
DB 223 ---AADDQLDANVMYALHF--YAGTHGQSLRDKANALSKAPIFTEWGTSDASGNGG 277
QY 263 --VDEA-TIMSSEQGVGLWMSKNGNGPEMEL---DLSMDAGNNLTANGNTIVNG 315
DB 278 VFIDQSR-----EMVNTLDSKNISWVWNLSDKQSSSALPGASKTGGWPLTDLTASGTYREN 337
QY 316 PYGLRETSR 324
DB 338 IRGTKDSSTK 346

```

RESULT 15

G69593
 cellulase (EC 3.2.1.4) bglc precursor - *Bacillus subtilis*
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence, revision 05-Dec-1997 #text, change 15-Oct-1999
 C:Accession: G69593; A26114; I40353; S24233; S49103; I39803
 R:Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabbre, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fume, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koester, P.; Kollingstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardino
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani
 A.; Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A.; Authors: Yoshikawa, H.F.; Zumsattel, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MID:98044033
 A:Accession: G69593
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-508 <KUN>

A:Cross-references: GB:299113; GB:AL009126; NID:92634090; PID:CA13696.1; PID:el183471.
 A:Experimental source: strain 168
 R:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranello, F.; Seld
 Nucleic Acids Res. 14, 9159-9170, 1986
 A>Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
 A:Reference number: A26114; MOID:87066783
 A:Accession: A26114
 A:Molecule type: DNA
 A:Residues: 10-508 <MAC>
 A:Experimental source: strain PAP15
 A>Note: part of this sequence, including the amino end of the mature form, was confirmed
 R:Lindahl, V.; Ae, K.; Tromsø, A.
 Antone Van Leeuwenhoek 66, 327-332, 1994
 A>Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C
 A:Reference number: I40353; MOID:95225656
 A:Accession: I40353
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-291 'N', 293-508 <LIN2>
 A:Cross-references: EMBL:X67044; NID:939776; PID:CAA47429.1; PID:939777
 R:Lindahl, V.; Ae, K.
 A:Cross-references: EMBL:X67044; NID:939776; PID:CAA47429.1; PID:939777
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S24239
 A:Accession: S24239
 A:Molecule type: DNA
 A:Residues: 10-291 'N', 293-508 <LIN2>
 A:Cross-references: EMBL:X67044; NID:939776; PID:CAA47429.1; PID:939777
 A:Experimental source: strain CK-2
 R:Wolf, M.; Geetzl, A.; Borriass, R.
 A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons
 A:Reference number: S49103
 A:Accession: S49103
 A:Molecule type: DNA
 A:Residues: 10-508 <MO>
 A:Cross-references: EMBL:Z29076; NID:9509266; PID:CAA82317.1; PID:9509267
 R:Seo, Y.S.; Lee, Y.H.; Park, U.H.; Kang, H.
 Korean J. Microbiol. 24, 236-242, 1986
 A>Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis
 A:Reference number: I39803
 A:Accession: I39803
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'A', 29-31, 'P', 33, 'PQ', 36
 A:Cross-references: GB:M38634; NID:9142657; PID:AAA22300.1; PID:9142658
 A:Experimental source: strain ATCC 6633
 A:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
 C:Genetics:
 A:Gene: bglC
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:39-508/Product: cellulase #status predicted <MAT>

Query Match 6.9%; Score 181; DB 2; Length 508;
 Best local similarity 19.3%; Pred. No. 5.8e-05;
 Matches 100; Conservative 102; Mismatches 188; Indels 128; Gaps 26;

QY 6 KRIFSTLSTLLASSTLF-----VSGSTANANSQ-FYVSGTLLDANGNPFVMSG 55
 DB 11 KRISIFITCLLTLLTGMGMIASPASAGTKIPVANKGOLSTKGTOLVNRDOKAVQLNG 70
 QY 56 I-NHGAHWYKQAT-TAIEGIANTGANTV---RIVLSDGGQWTKDDI-HTVNLISLAD 109
 DB 71 ISHGLQWYGEYVKNKSLKRLRDMGWITVFRAMYTADGGYIDNPSYKKNVKEAVEAKE 130
 QY 110 NHLVAPVEDATGYDSIASLNRADVWIMRSGALIGKEDTYIINTANEMFG--SWEGD- 166
 DB 131 LGIYIIVDWHILNDGNPNOKRERAKKEFEKEM--SSLYGNTPNVIYEIANEPNGVDYMKRDI 189

QY 167 -AWADGYKQAIPLRLNAGLNHTLWDAAGWGQFPQSTIHDIYGREVENADPO---RNTMFSI 222
 DB 190 KPYAE---EYISVIRKNDPDNIIV---GTGWSQDVND-----AADDLKDNANWTAL 237
 QY 223 HMEYAGNAGQVNTNDRVLNODLALVIEFGHRHNGD-----VDPA-TIMSSEQGV 277
 DB 238 HF--YAGTHGQFLDKANAYLSKAPIFVTEWGTSDASGNGVFLDOSREMLKYLDSKTI 295
 QY 278 GWLAWSRKNGPEWEYL---DLSDNAGNNLTAMGNTIVNGPYGLRETSR----- 324
 DB 296 SWVMNNLSDKQESSALKPGASKTGWRISDLSASGTFRNIIIGTIDSTKIDIPETPSKD 355
 QY 325 -----ISTYFTGGSGDGSPTTLYDFEGSMQGWGSSLSGGPMVTEMSSKSHSLK 377
 DB 356 KPTQENGISVYRAG-----DGSN-----NSNQIR 380
 QY 378 ADIOSSN-----SOHLYHVIQNTSLQNSRIQATVKAHMGWSVGNKMTARLYV- 425
 DB 381 POLQIKKNGNTYVDLKDYTAHYTKANKNG--QNP-----DDDYAQIGCGNVTYHKEFV 433
 QY 427 --KTGHGTYWY-----SGSEVPYINGSGTLLSLDLSN 456
 DB 432 LKPKQAGADTYLELGFKNKGTLP--GASTGNIQRLHN 467

Search completed: December 19, 2000, 17:06:42
 Job time: 1469 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:19:35 ; Search time 132.02 Seconds

(without alignments)
118.577 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNFGKRIISITSLILASS.....QSAQSDSSGQSTIYIDNVIYE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	737.5	28.3	383	1	MANA_STRLI
2	638	24.5	1331	1	MANB_CALSA
3	427	16.4	513	1	MANB_BACSM
4	234	9.0	448	1	MANB_CLOAB
5	229.5	8.8	400	1	GUN5_CLOAB
6	226.5	8.7	488	1	GUN5_BACAG
7	226	8.7	409	1	GUN1_BACSA
8	220.5	8.5	505	1	GUNN_ERMCA
9	219.5	8.4	444	1	GUNN_ERMCA
10	207	7.9	504	1	GUNN_ERMCA
11	188.5	7.2	499	1	GUN1_BACSU
12	181	6.9	499	1	GUN2_BACSU
13	180	6.9	466	1	GUN5_THERU
14	171	6.6	499	1	GUN3_BACSU
15	170.5	6.5	584	1	GUN2_CLOCE
16	168	6.4	800	1	GUN_BACSI
17	166	6.4	941	1	GUN_BACSI
18	165.5	6.3	459	1	GUNA_STRLI
19	164.5	6.3	825	1	GUN3_BACSA
20	160.5	6.2	510	1	YDYL_SCHRO
21	155	5.9	814	1	GUNE_CLODM
22	151.5	5.8	562	1	GUN1_ACTIE
23	147.5	5.7	429	1	GUNA_BUTRI
24	146.5	5.6	747	1	GUN2_CELFI
25	141	5.4	312	1	GUNA_RUMAL
26	140.5	5.4	426	1	GUN_BURSO
27	133	5.1	1039	1	AG43_ECCLI
28	132.5	5.1	1608	1	HXA_SERVA
29	132	5.1	438	1	EXG_CANAL
30	131	5.0	515	1	GUND_CLOLO
31	131	5.0	517	1	GUNA_CLOLO
32	129	4.9	566	1	GUN2_CLODM
33	129	4.9	2003	1	YDAB_ECCLI

ALIGNMENTS

34	128.5	4.9	890	1	BCN5_CLOPE	P08696 clostridium
35	127	4.9	426	1	GUN2_ERMCH	P07103 erwinia chr
36	123.5	4.7	645	1	LIP1_PROIU	P40601 photorhabdu
37	123.5	4.7	711	1	CDGT_BACST	P31797 bacillus st
38	122.5	4.7	1039	1	GUNB_CALSA	P10474 c endogluca
39	118.5	4.5	1829	1	FRPC_NEIMC	P51127 neisseria m
40	116.5	4.5	440	1	GUNB_CLOCL	P28621 clostridium
41	116.5	4.5	976	1	RTBP_ADDEB	O03553 bovine aden
42	115	4.4	1656	1	OMP_RICCA	O06653 r outer mem
43	114.5	4.4	865	1	HTRE_ECCLI	P31129 escherichia
44	114	4.4	802	1	PEFC_SALTY	P37868 salmonella
45	114	4.4	2334	1	WAPA_BACSU	Q07833 bacillus su

RESULT 1	MANA_STRLI	STANDARD:	PRT: 383 AA.
AC	P51529:		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	MANNA END-1,4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-MANNANASE) (1,4-BETA-D-MANNAN MANNAHYDROLASE).		
GN	MANA.		
OS	Streptomyces lividans.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
CC	Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.		
CC	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 36-42.		
RC	STRAIN-66 / 1326;		
RX	MEDLINE; 93207541.		
RA	Arcand N., Kluempfel D., Paradis F.W., Morosoli R., Shareck F.;		
RT	"Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence of the manA gene and characterization of the enzyme.";		
RL	Biochem. J. 290:857-863(1993).		
RN	[2]		
RP	REVISIONS TO C-TERMINUS.		
RC	STRAIN-66 / 1326;		
RA	Shareck F.;		
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.		
CC	- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES CELSIUS AND A PH OF 6.8.		
CC	- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS.		
CC	- SUBUNIT: MONOMER.		
CC	- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).		
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M92297; AAA26710.2; -.		
DR	INTERPRO; IPR001547; -.		
DR	PFAM; PF00150; cellulase; 1.		
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.		
KW	Hydrolase; Glycosidase; Signal.		
FT	SIGNAL		
FT	CHAIN		
FT	DOMAIN		
FT	SEQUENCE		
SO	383 AA; 39681 MW; SDB4B407C64E94C3 CRC64;		
Query Match	28.3%;	Score 737.5;	DB 1; Length 383;
Best Local Similarity	44.4%;	Pred. No. 1.8e-43;	

Matches 160; Conservative 56; Mismatches 121; Indels 23; Gaps 9;

QY 13 LSLALLSILFVSGTSTANANS---GPTYSGLTYLNGANGPFTMGINHGANYKQDATT 69
DB 19 LGLLFA---LAGPSGRAEAAGIHWNSNGRVGNSAFVWRGVNNAHYTYVDR-TG 72
QY 70 AIEGANTGANTVRIYLDGCGMTKDDIHTVRLISLAEDNHLVAPEVDATG--DSEI 127
DB 73 SIMDIAGKANTVRYVLSGGRKTKTSASVSLIQCCKANKYICVLEHDTTGCKDA 132
QY 128 ASLNRAVDVIERKSL-IGKEDVTIINANEMFGSMGEGDAMADYKQAIPLRLNAGLNH 186
DB 133 TSLDQAGDYVGVKSAAMRAQEDYVYVINGENEFNTNAAMTDATKSAIGLRGAGLGH 192
QY 187 TLMVDAAGCO-FPOSIDHVGREVNADPQRTMFSIHMTETVAGNASOVRTIDRYLNO 245
DB 193 ALMVAIDPNMGQDWSGTMRNNAVSFSDPRNTVFIFHMG-VYDPAEVRDLNLFVN 251
QY 246 DLALVIGFGRHTNGDVDEATIMSYSEORGVLAMWKGNGPEVEYLDLSNDAGNLT 305
DB 252 GLPIVVGEGDQSDNDPDEDAIMATQASLGVOYLGNMSNGGVYELDMVNGPDPNSL 311
QY 306 TANGNTIVGPGYGLRTSRSLTYF-TGGSDDGTSPT-----TLVDFEGSMOGWTGS 356
DB 312 TSMGNRLYSGNIASTSTATVYGGGSGTGTAENGYPYCVNGASDPDGDGMGEMENS 371

RESULT 2
MANB-CALSA STANDARD; PRT; 1331 AA.
ID MANB-CALSA
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BETA-MANNANASE/ENDOGALUCANASE A PRECURSOR [INCLUDES: MANNAN ENDO-1,4-
DE BETA-MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-
DE MANNANASE) ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULOSE)].
GN MANB.
OS Caldocellum saccharolyticum (Caldicellulosigranulosus saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermomicrobacter group; Caldicellulosigranulosus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93119139.
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
RT multidomain enzyme";
RL Appl. Environ. Microbiol. 58:3864-3867 (1992).
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE; 91247819.
RA Luehl E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'";
RL Appl. Environ. Microbiol. 57:694-700 (1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT
CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC
CC LINKAGES IN MANNANS, GALACTOMANNANS, AND
CC GALACTOGLUCOMANNANS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
CC CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>
or send an email to license@ebi.ac.uk).

CC EMBL; L01257; AAA71887.1; -;
CC EMBL; M36063; AAA72861.1; -;
CC PIR; B43745; B43745.
CC PIR; A48954; A48954.
CC HSSP; Q06851; INBC.
CC INTERPRO: IPR001547; -;
CC INTERPRO: IPR001956; -;
CC PRAM; PF00942; CBD_3; 2.
CC PFAM; PF00150; cellulase, 1.
CC DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
CC DR Hydroxylase; Glycosylase; Cellulose degradation; Signal;
CC Multifunctional; enzyme.
CC SIGNAL 41
CC CHAIN 42; 1331
CC DOMAIN 42; 325
CC DOMAIN 326; 361
CC DOMAIN 362; 518
CC DOMAIN 519; 564
CC DOMAIN 565; 720
CC DOMAIN 721; 780
CC DOMAIN 781; 1331
CC ACT_SITE 162; 162
CC ACT_SITE 257; 257
CC CONFLICT 338; 338
CC CONFLICT 340; 346
CC TPTPTPT -> ROHORO (IN REF. 2).
CC SEQUENCE 1331 AA; 146892 MW; FPCAS1BB8DBF0E0 CRC64;

Query Match 24.5%; Score 638; DB 1; Length 1331;
Best Local Similarity 31.4%; Pred. No. 5,9e-36;
Matches 166; Conservative 84; Mismatches 208; Indels 70; Gaps 14;

QY 6 KRITSITSLASISLIFS-----GTSTANANGFYVSGTLYDANGNPFVNRGIN 57
DB 9 KRWLSVLTCTVFLNITLFIANVTILPKVGAATSN-----DGVKKIDTS---TLIGTN 57
QY 58 HGHMYKQDQATEGANTGANTVRIYLDGCGMTKDDIHTVRLISLAEDNHLVA-VP 126
DB 58 HAHCTYKRDLPALRGIRSGMANSVRYVLSNGIYWTIIPSEVANTISLSIGFAIIL 117
QY 117 EVHDATG--DSIASINRAVDVIERKSLIGKEDVTIINANEMFGSMGEGDAMADYK 173
DB 118 EVHDTTGGEDGACSLAQAVEYKEIKSVLDGDEEVIIINGEPPNNNYQWVNDTK 177
QY 174 QAIRLRNAGNHLTLPDAAGWQ-FPOSIDHVGREVNADPQRTMFSIHMTETVAGNA 232
DB 178 NAIKRLNAGKHTIMVADAPMGQDWSMTWRKDNQASIMEDPPLNLFVSIHMVG-VYNTA 235
QY 233 SOVRTNIDRYVNDOLATVIGFGRHTNGDVDEATIMSYSEORGVLAMWKGNGPEVE 292
DB 237 SKVEEYIKSFYDKPLPLYIGFQHGTGDPDEALRYAKQIKGLFSSWCGNSSYVG 296
QY 293 YLDSNDWAGNNLTWANGNTIVNGPYGLRTSRSLTYFTGGSDDGT-SPTLLYDFEGSMQ 351
DB 297 YLDVNNMWDPNNPYPGWQWKTNAIGTSPPTPTPTPTPTPTPTPTPTPTPTPTPTPT 356
QY 352 GWTSSISLG-----GPAVAVWSSKSGSHSLKADITQSSNOHYLVT 393
DB 357 PVSTPATSGQIKVLYANKETNSTNTIRPWL-----KVANSSSSIDLSRYIRIYTYV 410
QY 394 QNTSLQNSRLOATVKANNGSVNGMTARLYKTH-----GTVWSSGSPVPI 442
DB 411 -----DGRQNSAI--SDNAQIGASVNTKFKYKLSVSGADYIETIGKSGAGQLPG 462
QY 443 NGSSGTTSLDSNVONTLQVREIG-VQFQASDSSGQTSYIDNVY 489
DB 463 KDTGEIGMRFKIDWMSVNGNDMSWISQTSYGENEKYATYIDGVLY 510

	RESULT	3
ID	MANN_BACSM	STANDARD; PRT: 513 AA.
AC	P16699;	
DT	01-AUG-1990 (Rel. 15, Created)	
DT	01-AUG-1990 (Rel. 15, Last sequence update)	
DT	01-FEB-1991 (Rel. 17, Last annotation update)	
DE	MANNAN ENDO-1,4-BETA-MANNOSIDASE A AND B PRECURSOR (EC 3.2.1.78)	
DE	(BETA-MANNANASE) (ENDO-1,4-MANNANASE).	
OC	Bacillus sp. (strain AM-001).	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/staphylococcus group; Bacillus.	
RN	[1]	
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RX	MEDLINE: 90146329.	
RA	Akino T., Kato C., Horikoshi K.;	
RT	"Two Bacillus beta-mannanases having different COOH termini are	
RT	produced in Escherichia coli carrying pMAN5."	
RL	Appl. Environ. Microbiol. 55:3178-3183(1989).	
CC	-1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC	
CC	LINKAGES IN MANNANS. GALACTOMANNANS, GLUCOMANNANS, AND	
CC	GALACTOGLUCOMANNANS.	
CC	-1- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.	
CC	--	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/	
CC	or send an email to license@sdb-stb.ch).	
CC	--	
DR	EMBL: M31797; AAA2586.1; -	
DR	PIR: A37219; A37219.	
DR	INTERPRO: IPR000805; -	
DR	PRINTS: PRO0739; GLHYDRLAS26.	
KW	Hydrolase; Glycosidase; Signal.	
FT	SIGNAL 1 26	
FT	CHAIN 27 513 MANNAN ENDO-1,4-BETA-MANNOSIDASE A.	
FT	CHAIN 27 365 MANNAN ENDO-1,4-BETA-MANNOSIDASE B.	
SO	SEQUENCE 513 AA; 58430 MW; 8BD105F622CDB5A8 CRC64;	
<hr/>		
Query Match	16.4%; Score 427; DB 1; Length 513;	
Best Local Similarity	26.9%; Pred. No. 4.3e-22;	
Matches 149; Conservative 76; Mismatches 215; Indels 114; Gaps 23		
QY	5 FKIKFSITSLDLASSIFVSGSTSYANASGYYVGTTLYDANGPFFVM-----	53
Db	4 YKVAVFVAAFIMEFESVLPTISMSEARN-----GALLSNNAQTNTANYSWLANLPN	55
OY	54 ---RGINHG-AMYKDAQTATAGIANTGANTVRIVLS-DGGQW-----TKDDHTVRN-	102
Db	56 KSKNRVVVGSHGEGYSDSTLAWKQCARELTGKMPGILLSCDYKNMGQRLVAVDAQISYGCNQ	115
OY	103 -LISTAEONHLYAV-----PEYDHATGYDSIASLNBRANDY-----WIEKRSALI	145
Db	116 ELINFWNOGGLVTYSIVHMNPFGSHGEGENTKLPTPSQFNLTNNHRTTEGRRWKMMDKMA	175
OY	146 GKED-----TVIINIANEMFGSWGDM--ADGYKAQIPRLRNAGLN-----H	186
Db	176 DGDLDLQNNGTVLFRLPLHEMNGEM---PWMAEGEINDQDQRANKYISAKMDMTQYFIH	232
OY	187 TLAVDAAGWGQFPQSIIHDYGREVENADPOKNTMFSSIHMYETAGNASQVRYTNIDRYLAND	246
Db	233 ERKLNLNLIMVYSPDYRDH---VTSYYPGANVYDIYALDSY-HPDHSITDQYNRMIALD	288
OY	247 LALVIGEIGHRT-NGVDDEATIMSYSQ---RGVGNLAMS--WR--GNGPFWEYTLDSL	298
Db	289 KPFAFAETIGPPESMAGSFEYSNTAIKQIKYTRTYVFLVAMNDKWSPHNRRGAM--DLFN	345

[illegible]

```

DB 9 NEFKTFESFLAVM---MLFTVLGTNTYKAPAAFTSREGGQLKVVGSQLCDSNGKPIQLK 65
OY 55 GI-NHGHAMV-----KDOATTAIEGIANTANTVRIVL--SDGQMT--KDDI 97
DB 66 GMSHSLQVYVFNVDMSKFLRDK-----MGVAVIAAMVTNNGYISNPSQK 115
OY 98 HVTNRILISIAEDNHLVAPEVDAGYDSIASLNRAVDWIMERSALICEKEDTVIINLAN 157
DB 116 EIKIKIVQDAIDLMNVIIIDWHILSDNPNPTYEKAKSFFOEAEH-YGKYSNVYEICN 174
OY 158 EMFGSMEGAMADGYK---QAIPLRNAGLNHTLWDAAGQFQSHDYGREVENAD 213
DB 175 EPNNG---GNWANDIKPYNTIIPAIRADPNNTIIVGISTMSQVDIAD-----N 223
OY 214 POR--NTMFSIHMYEYAGNASQVTRNIDRVLNODALVIGEGHRTNGVDDEATINSY 271
DB 224 PLKYSNIMTCHF--YAGHHTQSLRDKINVAMSKGAIPEVTMGTSADAGN----- 272
OY 272 SEORGVGWLAMSKWKGPEWEYIDLSNDV---AGNNLWAGNTLVNGYGLRETSRLST 327
DB 273 -----GGP---YIDESQKWVDFWASKNI-SWTN-----WALCDKSEASA 307
OY 328 VFTGGSDGTSPTLTLYDFEGSMQWGTGSLG-----GPPVAVTEWSSKSGSLK 377
DB 308 ALKSGSSTTG-----GNTDSDLTSSGLFVKKSIG-----SNTTSQTS 346
OY 378 ADIQLSSNQHYLVNTQNTSLQONSRIQATVKHANNGSVNGMTARLYK--TGHGYTW 434
DB 347 PFTSLQGTYSAQVTLNSDNDVSIHYTTDGTPTSSPVTYSPITISKTTVAFTT 406
OY 435 YGSGFPVINGSGTSLDLSNQNINSQV 463
DB 407 KTG---MTDSNITSAYVTITSTNDPVKQV 431

```

RESULT 5

```

GUN5_BACAG STANDARD: PRT; 400 AA.
ID GUN5_BACAG
AC 085465;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENDOGLUCANASE 5A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (ALKALINE
  CELULASE).
GN CEL5A.
OS Bacillus agaradhaerens (Bacillus agaradhaerens).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 8721;
RA Bjornvad M.E.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329.
RC STRAIN-AC13 / NCIMB 40482;
RX MEDLINE; 98153671.
RA Davies G.J., Dauteur M., Brzozowski A.M., Bjornvad M.E.,
  Andersen K.V., Schuelein M.;
RT "Structure of the Bacillus agaradhaerens family 5 endoglucanase at 1.6-
  A and its cellobiose complex at 2.0-A resolution."
RL Biochemistry 37:1926-1935(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329.
RC STRAIN-AC13 / NCIMB 40482;
RX MEDLINE; 98384136.
RA Davies G.J., Mackenzie L.F., Varrot A., Dauteur M., Brzozowski A.M.,
  Schuelein M., Withers S.G.;
RT "Snapshots along an enzymatic reaction coordinate: analysis of a
  retaining beta-glycoside hydrolase."
RL Biochemistry 37:11707-11713(1998).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
  LINKAGES IN CELLULOSE.

```

```

CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
  HYDROLASES).
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
  or send an email to license@isb-sib.ch).
CC
DR EMBL; AF067428; AAC19169.1; -.
DR PDB; 1A3H; 16-MAR-99.
DR PDB; 2A3H; 16-MAR-99.
DR PDB; 3A3H; 16-MAR-99.
DR INTERPRO; IPR001547; -.
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACT_SITE 165..165 PROTON DONOR.
FT ACT_SITE 254..254 NUCLEOPHILE.
SQ SEQUENCE 400 AA; 44702 MW; 3F9C66F9B3C6FEF CRC64;

```

Query Match 8.8%; Score 229.5; DB 1; Length 400;
 Best Local Similarity 23.4%; Pred. No. 9.5e-09;
 Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

```

OY 6 KRIFSTLSLASSILFVSGSTANANS-----GYVSGTLLYDANGNPFVNGI-NH 58
DB 2 KIIITLIVLMTALFISGNTADNDVYEEHQSLISNGELVNERGEVQLKGMSSH 61
OY 59 GHMT-----KDOATTAIEGIANTANTVRIVL--SDGQMTKDDI-HTVANL 103
DB 62 GLQWYGFQFVNYESMKLRD-----DWGINFRAMYTSSGXYIDDPSEYKEVGA 111
OY 104 ISLADNHLVAVPEYHDANGVDSIASLNRAVDWIMERSALICEKEDTVIINANFEGSM 163
DB 112 VVAALIDLDIYIIDWHILSDNPNPTYEKAKDFDEM-SELKGYDPNVIYELANEPNS- 169
OY 164 ESDAMADGYK---QAIPLRNAGLNHTLWDAAGQFQSHDYGREVENADPQRTM 219
DB 170 -DYTMGNQIKPYAEVYIPIRNDPNNTIIV---GTGWSQDVH--HAADNOLADP--NVM 222
OY 220 ESIHMYEYAGNASQVTRNIDRVLNODALVIGEGHRTNGVDDEATINSYSEQ 274
DB 223 YAFHF--YAGTIGQNLROQVDYALDQGAIFVSEWGTSAATGDGVFLDEQVWIDFME 280
OY 275 RGVGWLANS-----WKGPEWEYIDLSNDMAGNNLTANGNTLVNGYGLRETS 323
DB 281 KNLNMANNSLITKDESSAALMGANP-----TGGWTEALSPSGTVRE--KIRESA 330
OY 324 RLSTVFTGGSDGTSPTLTLYDFEGSMQWGTGSL-----SGPPVAVTEWSS 369
DB 331 SIIPSDPTPSDPEPDPTPSDPEYPAWDPNQIYTNIEIYHNGQLQAKWMT 384

```

RESULT 6

```

GUN1_BAC4 STANDARD: PRT; 488 AA.
ID GUN1_BAC4
AC P06566;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
GN CELA.
OS Bacillus sp. (strain N-4).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87056924.

```

RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RL *Bacillus* sp. strain N-4 and their strong homology.";
 CC J. Bacteriol. 168:479-485(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14781; AAA22301.1; -
 CC PIR: A25156; A25156.
 CC INTERPRO: IPR001547; -
 CC PFM: PFM0150; cellulase: 1.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase.
 CC ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
 CC ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
 CC SEQUENCE 488 AA; 54264 MW; 972482453D283B5 CRC64;
 SQ
 Query Match 8.7%; Score 226.5; DB 1; Length 488;
 Best Local Similarity 23.1%; Pred. No. 2e-08;
 Matches 98; Conservative 67; Mismatches 157; Indels 103; Gaps 22;
 QY 6 KRIEFTITSLASSILFVSTGTSTAN-----ANSGEVYSTTLYDANGNPFVARGI-NH 58
 DB 2 KLTITLFIYFLTLA--LLEFGNSTSANNGSVYNGQLSIOGQLVYEHGDPVLQKMSH 59
 QY 59 GHAWKQDATTAEIGANTGANTVRIYVSDGGQWTKD-----IHT----- 99
 DB 60 GLQWYG-----QFVNDSTIKWLRDMDGITVFRAMTSSGGYIEDPS 101
 QY 100 ----VRNLISLAEDNHLVAVPEVDATGYDSIASLNRAVDYIEMRSALIGKEDYIINI 155
 DB 102 VKEKKEAVEAIDIGIYIIDMHILSDNDPNIIYKEAKEFEFDEM-SALXGDPVNIYI 160
 QY 156 ANEWFG---SWEGDAWADGT-KOAIPLRNAGLNHTLMDVDAAGWQFPOSIDYGREVN 211
 DB 161 ANEPGNHNVNR--DSHIKRYAEVPIPIRANPNNIYIVGTATWS---QDVHEADNQD 215
 QY 212 ADPQNTMFSIMHYEYAGNASOVRTNDRVYNODLALVIGFGRHTNGD-----VDEN 267
 DB 216 -DP--NVMYAFHF--YAGTHGOQLRQVDYALSRGAALFVSEMGTSAAATGDGVFLDE 270
 QY 268 I-MSYSEQRGVGLAWS-----WKNGPEWEYIDLSDNAGNLTLMANGTIYNG 315
 DB 271 VMDIDMDEKRLSWANWSLTHKDESSAALMPGANP-----TGGWTAALSLSGAFVRE- 322
 QY 316 PYGLRETSRLSTVFTGGSD---GGTSPTLYLDFEGSMQWGTSSS-----SGGPPA 364
 DB 323 --KIREASITPPSDPTPPSDPGEPPDPFPSPD-PGEYAMPNPQIYTEIYHNGQLNQ 379
 QY 365 VTEM 369
 DB 380 AKWMT 384
 RESULT 7
 GUN2_BACSA STANDARD; PRT; 409 AA.
 AC P06565;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE).

GN CELL.
 OS *Bacillus* sp. (strain N-4).
 CC Bacteria, Firmicutes; *Bacillus*/Clostridium group;
 CC *Bacillus*/Staphylococcus group; *Bacillus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87056924.
 RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RL *Bacillus* sp. strain N-4 and their strong homology.";
 CC J. Bacteriol. 168:479-485(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14729; AAA22299.1; -
 CC PIR: B25156; B25156.
 CC INTERPRO: IPR001547; -
 CC PFM: PFM0150; cellulase: 1.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase.
 CC ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
 CC ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
 CC SEQUENCE 409 AA; 45690 MW; 98886606B6BAD3F CRC64;
 SQ
 Query Match 8.7%; Score 226; DB 1; Length 409;
 Best Local Similarity 23.2%; Pred. No. 1.7e-08;
 Matches 104; Conservative 62; Mismatches 147; Indels 136; Gaps 2;
 QY 6 KRIEFTITSLASSILFVSTGTSTA-----NANSGEVYSTTLYDANGNPFVARGI-N 57
 DB 2 KLTITLFIYFLTLA-LFIIGNTAADDYVYEEHGOISNGELVNDRGEPVQKMS 60
 QY 58 GHAWKQDATTAEIGANTGANTVRIYVSDGGQWTKD-----IHT----- 99
 DB 61 HGLQWYG-----QFVYESMKWLRDMDGITVFRAMTSSGGYIEDP 102
 QY 100 ----VRNLISLAEDNHLVAVPEVDATGYDSIASLNRAVDYIEMRSALIGKEDYIINI 154
 DB 103 SVKEKKEAVEAIDIGIYIIDMHILSDNDPNIIYKEEAKDEFDEM-SELYGDYENVIYE 161
 QY 155 IANEWFGS--WEGD--AWADGYKOAIPLRNAGLNHTLMDVDAAGWQFPOSIDYGREV 209
 DB 162 IANEPGSDVTDNDQIKPAE--EVIPIYRNNDNNIITV--GTGTMSQDVH-HAADN 214
 QY 210 ENADPQNTMFSIMHYEYAGNASOVRTNDRVYNODLALVIGFGRHTNGD---VDE 265
 DB 215 QLTDP--NVMYAFHF--YAGTHGOQLRQVDYALDGAALFVSEMGTSAAATGDGVFLDE 270
 QY 266 ATI-MSYSEQRGVGLAWS-----WKNGPEWEYIDLSDNAGNLTLMANGTIY 313
 DB 271 AQVWIDFMDERLWSANWSLTHKDESSAALMPGASP-----TGGWTAALSLSGAFV- 322
 QY 314 NGPYGLRETSRLSTVFTGG-----SDGGT-----SPTLYD 345
 DB 323 -----REKIRSATTPSDPTPPSDPGEPPDPFPSPDPTPPSDPQDYRAMDENTIT 376
 QY 346 FE-----GSM---QGWTSLSG---GPW 363
 DB 377 DEIVYHNGQLQAKWMTQNOEPGDYGPW 405
 RESULT 8

Query Match 8.4%, Score 219.5; DB 1; Length 444;
Best Local Similarity 21.7%, Pred. No. 5.2e-08;
Matches 108; Conservative 71; Mismatches 193; Indels 125; Gaps 21;

2 NNGFKIFSTLTLASSILFVSGSTANANS-
6 NOYAKKLGVTTVGLGMSLSFSAISATPETHGOLSTINGLVDPQKRVOLRGVSSHG
60 HAWYKQATTAEGIANGTANTRIYVSDGGQWTRD
66 LQWFGD
97 IHTVNLISLAEDNHLVAVPEVHDATGIDSLNRAVDYIEMKSAIIGKEDYIINIA
108 ANKVEAAVAAAGSLGVYIIIDHILSDNDPNYKAQAKFFEM-AGLYGSSPNVYIETA
157 NEMFG-SWEGD-AMADGYKQALPRLNAGLNHTLVADAAGWGPPOSIDHGREVFA
167 NEPNQGVTHWGOIRPYA-LEVDTIRSKDPDNIIV-GTGWSQDIDH-AAQNO
213 DQORTMSIHMYEYAGNASQVNTIDRVLNQDLALYIGFGRHTNGD-VDEAT
218 LPDPRTLYALH-YAGTHGQFLRDIRDYAQRGAIFVSEMGSDASGNGPFLPESQT
268 IMYSIQGVGHLANSWKGNGPEWEYL-DLSNDAGNNITANGNTIYNGPYGLRETS
276 WIDFLNNGVSWVWNSLSDKSEASALAPGASKSGWTEQNLSTGKFEV-REQI
324 RUSTYFGGSGGTPTLYDFEGSMQWGTSSLSGPMATWEMSKSHSLKADIOLS
329 R-AGANTGGGTPPT-PTTEPTNP-NGTTGDVVL-
384 SNSQHLVIONTSIQOQSRIGATYKHAMGVS-NGMTARLYKTHGTYTSGSEV
383 -QYRVNDNPS-DDAIRMAFVNINKTGSTPIKLSQVRYFHD-
441 PINGSGTTLSDLSNV 457
404 -DGRGANLFFDMANV 418

RESULT 10
GNNL_ERMCA STANDARD; PRT; 504 AA.
AC 059395;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ENDOGLUCANASE VI PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE VI)
DE (CELLULOSE VI).
GN CELV1.
OS *Erwinia carotovora*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SCC3193;
RX MEDLINE; 95231512.
RA Mae A., Helsinki R., Palva E.T.;
RT "Structure and regulation of the *Erwinia carotovora* subspecies
RT *carotovora* SCC3193 cellulase gene *celv1* and the role of cellulase in
RT phytopathogenicity";
RL Mol. Genet. 247:17-26(1995).
CC -I- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in its
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: X79241; CAA55823.1; -
DR HSSP: 006851; INBC
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR001956; -
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosylase; signal.
FT SIGNAL 1 31
FT CHAIN 32 504
FT DOMAIN 32 334
FT DOMAIN 335 352
FT DOMAIN 353 504
FT ACT_SITE 168 168
FT ACT_SITE 256 256
SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 7.9%, Score 207; DB 1; Length 504;
Best Local Similarity 22.3%, Pred. No. 4.4e-07;
Matches 110; Conservative 63; Mismatches 191; Indels 130; Gaps 22;

6 KRIEFTLSLTLASSILFVSGSTANANS-
10 RULTLGVTTMGLMSLSFSAISATPETHGOLSTINGLVDPQKRVOLRGVSSNGLOV
64 KDQATTAEGIANGTANTRIYVSDGGQWTKD-IHTVNLISLAEDNHLVAVPEV-
70 GD-
119 -HDAITYDSTASLNRAVDYIEMKSAIIGKEDYIINIANM
111 VKEAVAAAGSLGVYIIIDHILSDNDPNYKAQAKFFEM-AGLYGSSPNVYIETA
160 FGS-SWEGD-AMADGYKQALPRLNAGLNHTLVADAAGWGPPOSIDHGREVFA
170 NSQVTHWGOIRPYA-LEVDTIRSKDPDNIIVGSGTW-
216 RNTMFSIHMYEYAGNASQVNTIDRVLNQDLALYIGFGRHTNGD-VDEATIS
221 PNTLYALH-YAGTHGQFLRDIRDYAQRGAIFVSEMGSDASGNGPFLPESQT
271 YEDQGVGHLANSWKGNGPEWEYL-DLSNDAGNNITANGNTIYNGPYGLRETS
279 FLNNGGISVWNSLSDKSETSAALVAGASKSGWTEQNLSTGKFEV-REQIR-
327 TFFTGSGSGGTPTLYDFEGSMQWGTSSLSGPMATWEMSKSHSLKADIOLSSNS
330 -AGAGLSGGTPPTMPPEPTNPNGTIG-DIYL-
387 OHYLVIONTSIQOQSRIGATYKHAMGVS-NGMTARLYKTHGTYTSGSFVPIN
360 -QYRVNDNPS-DDAIRMAFVNINKTGSTPIKLSQVRYFHD-
444 GSSGTTLSLSDLSNV 457
402 GRGANLFFDMANV 415

RESULT 11
GNNL_BACSTU STANDARD; PRT; 499 AA.
AC P07983;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).

GN BGIC OR GID.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DLG:
 RX MEDLINE: 87194581.
 RA Robson L.M., Chambliss G.H.;
 RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
 RL J. Bacteriol. 169:2017-2023(1987).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M16185; AAA2496.1; ALT_INIT.
 DR PIR: A26874; A26874.
 DR HSSP: 006851; INBC.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR001956; -.
 DR PFAM: PF00942; CBD_3; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 499
 FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 SO SEQUENCE 499 AA; 55187 MW; 339D04EE95A63EE1 CRC64;

Query Match 7.2%; Score 188.5; DB 1; Length 499;
 Best Local Similarity 21.1%; Pred. No. 8e-06;
 Matches 78; Conservative 75; Mismatches 133; Indels 83; Gaps 17;

6 KRIFITISLASSILF-----VSGSTANANSQ-FYVSGTTLIDANGNPFVWNG 55
 2 KRISIFITCLIAVLITMGGLPSPASAGTKTPVAKNGOLSKIGTQLVNRDGNKAVOLNG 61
 56 I-NHGSAWKQDQATTAIEGANTGANTVRIYVSDGGQWTKD-----IHT----- 99
 62 ISSHLOWTGD-----FVKNDISKALRDWMGIVTFAAAYTADGGYI 103
 100 -----VRNLSLAEDNHLVAPEVDHATGYDSIASINRAVDYWIERSALIGKEDTV 151
 104 DNPVYKNAKKEAVEAKKEIGIYIIDMHILNDGNPNQNEKKEKFEK-SSLYGTPTNV 162
 152 IININEMNG--SWEGD--AMADGYKQATPRLRNAGINTLAVDAAGWQFQSHDYR 207
 163 IYEINENPDVNMKRDIKPIAE---EVSIVRKNDPDNITIV---GGETWSQDVND--- 213
 208 EVFNADPQ---NNTFSSIMYEYAGGNASQVRTNIDRVNODLAVIGEGRRHTNGD-- 262
 214 ---AADDQKADNVMYALHF--YAGTHGSLDKXKANYALSKAPFIVTWMGSDASGNG 268
 263 --VDEA-TIMSTSEORGWGLAMSKNGKPEWEYL---DLSNDNAGNNLTAMGNTIYNG 315
 269 VFLDQSRKLTALDSKNISWVMNLSDKQSSSALKPKASKGTGWPPLDTLTSAGTFVREN 328
 316 PYGLRETSR 324
 329 INGTQDSTR 337

RESULT 12
 GN2_BACSU STANDARD; PRT; 499 AA.
 AC P10475;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 GN DE (CARBOXYMETHYL-CELLULOSE) (CMC-ASE) (CELLULOSE).
 OS BGIC OR GID OR EBL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PAP15;
 RX MEDLINE: 87066783.
 RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
 RA Moraneli F., Seligy V.;
 RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
 RL Nucleic Acids Res. 14:9159-9170(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CK-2;
 RX MEDLINE: 95225656.
 RA Lindahl V., Aa K., Tronsmo A.;
 RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
 RT subtilis CK-2.";
 RL Antoine Van Leeuwenhoek 66:327-332(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Rose M., Ertlan K.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 30-45.
 RC STRAIN-CK-2;
 RX MEDLINE: 95225655.
 RA Aa K., Flensburg R., Lindahl V., Tronsmo A.;
 RT "Characterization of production and enzyme properties of an
 RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
 RT compost soil.";
 RL Antoine Van Leeuwenhoek 66:319-326(1994).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z29076; CA882317.1; -.
 DR EMBL: X04689; CA28592.1; -.
 DR EMBL: X67044; CA47429.1; -.
 DR EMBL: Z73234; CA497610.1; ALT_INIT.
 DR EMBL: Z99113; CAB13696.1; ALT_INIT.
 DR PIR: A26114; A26114.
 DR HSSP: 006851; INBC.
 DR SUBTILIS: BG10437; BGIC.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR001956; -.
 DR PFAM: PF00942; CBD_3; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 499
 ENDOGLUCANASE.

FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 FT CONFLICT 283 283 S -> N (IN REF. 2).
 SQ SEQUENCE 499 AA; 55287 MW; 8F735FF11B3EAE2 CRC64;

Query Match 6.9%; Score 181; DB 1; Length 499;
 Best Local Similarity 19.3%; Pred. No. 2.6e-05;
 Matches 100; Conservative 102; Mismatches 188; Indels 128; Gaps 26;

QY 6 KKFTSLSLASSILF-----VSGSTANANS-G-FYVSGTLLYANGPFWARG 55
 DB 2 KRISITITLTLTLTMGMGMIASPSASACTKTPVANKGSLSTGTLVNRDGAVALKG 61
 QY 56 I-NHGANWYKQAT-TAIEGIANGTANTV--RIVSDGGQWTKDDI-HTVNRLISLAED 109
 DB 62 ISSHGLQWYGEYVNDKSLKFLRDDMGITVPRAMYTADGGYIDNPVKNVKEAVEAKE 121
 QY 110 NMLVAVPEVDATGYDSIASLNAVVDWITEMRSALIGKETVYIINTANEMFG--SWEGD- 166
 DB 122 LGIYIIVDMLNDGNPNQKKEKFEKEM--SSLGNTPTNVIYELANENGDNVNRDI 180
 QY 167 -ANADGYKQAIPLRLNAGLNHTLAVDAAGQGFPOSIDHYGREVENADPO---RTMFESI 222
 DB 181 KPYAE---EYIVIRKNDPNIITV---GTGTSQDVND-----AADDLKANVATL 228
 QY 223 HMYEYAGNASSQVNTINDRYLNODLALVIGFGRHRTNGD---VDEA-TIMSYSEQGV 277
 DB 229 HF--YAGTHQOFLRDKANVALSKGAPLFYTEMGSDASGQVFLDQSRMLKTLKIDSKI 286
 QY 278 GWLMSKMGKPEWEYL---DLSNDAGNNLWANGTINYNGRYGLAETSR----- 324
 DB 287 SWVNMMLSDKQESSALKPKGASRTGWRSLDLSASGTFVENILGTQDSTKDIPEPSSKD 346
 QY 325 -----LSTVFTGGSGDGTSPITLYDFEGSMOGWTGSSLSGGPNAVTEMSSGSHLK 377
 DB 347 KPYQENGISVQYRAG-----DGSN-----NSQIR 371
 QY 378 ADIQLSNN-----SQHYLVYQNTSLQONSRIQATVKRANMGVNGMTARLYV- 426
 DB 372 POLQIKNGMTYVDLKDVTARYWKAKNKG--QNF-----DQDYAQICGNVTHKFTV 422
 QY 427 --KTGHGYTY-----SGSFYVINGSSGTTLSLDSN 456
 DB 423 LHRKQGADTYLELGFKNGLIAP--GASTGNIQLRLN 458

RESULT 13
 GUN5_THEFU STANDARD; PRT; 466 AA.
 ID GUN5_THEFU
 AC 001786;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4) (CELLULOSE E-5) (CELLULOSE E5).
 GN CELE.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TX;
 RX MEDLINE; 91258320.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RT "DNA sequences of three beta-1,4-endoglucanase genes from
 Thermomonospora fusca".
 RL J. Bacteriol. 173:3397-3407(1991).
 RP REVISIONS.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE OF 137-142 AND 157-166.
 RC STRAIN-TX;
 RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
 RT "Activity studies of eight purified cellulases: specificity,
 RT synergism, and binding domain effects".
 RL Biotechnol. Bioeng. 42:1002-1013(1993).

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; L01577; AAC09379.1; -.
 DR PIR; C42360; C42360.
 DR HSSP; P07986; 1EXH.
 DR INTERPRO; IPR001547; -.
 DR INTERPRO; IPR001919; -.
 DR PFAM; PF00553; CBD_2; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT CHAIN 1 36
 FT ACT_SITE 37 466 ENDOGLUCANASE E-5.
 FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 466 AA; 49800 MW; 1CF0ADBEFF2DEF82E CRC64;

Query Match 6.9%; Score 180; DB 1; Length 466;
 Best Local Similarity 25.0%; Pred. No. 2.8e-05;

Matches 82; Conservative 49; Mismatches 129; Indels 68; Gaps 19;

QY 26 GSTANANS-G-FYVSGTLLDANGNPFVNRGIN-HGHWYKQDAT-TAIEGIA-NTGANT 81
 DB 160 GTGTPEYERIKVQVCGTQCDDEHGNPVQLKGMSTHGIQWDDHCLTDSLDALADWKADI 219
 QY 82 VRI---VLSDGGQWT---KDDIHVYRNLISLAEDNHLVAVPEVDATGYDSIASLNRAY 134
 DB 220 IRLSMYIQEDGYEINPRGFTDRH---QLIDMATARGLYIVDWHILTPDPHYNIDRAK 276
 QY 135 DYIEMRSALIGKEDVYIINANEMFGSWEGDAMAD--GY-KQAIPLRLNAGLNHTLAVD 191
 DB 277 TFFAEIAORHASKTN-VLAEIANE---PNGVSWASIKSYAEVIVPIRDRDPSVYIVG 331
 QY 192 AAGMGQFPOSIDHYGREVEVADP--QRTMFSIHMYEYAGNASSQVNTINDRYLNODLAL 249
 DB 332 TRKMSLGVG--EESGFAELIANPVYASNTATAHFY-----AASHRDYLNALREASEL 364
 QY 250 ---VIGFGRHRTNGD-----VDEATISYSEORGVGLANSWK----- 285
 DB 385 FPFVTFEFGRETYTGOANDFOVADR--YIDLAERKIGTKNNYSDDFSGAVFOPGTC 442
 QY 286 -GNGPEWEYIDLNDWAGNLTWAGNTI 312
 DB 443 ASGGP-----WSSSLKASGQWV 460

RESULT 14
 GUN3_BACST STANDARD; PRT; 499 AA.
 ID GUN3_BACST
 AC P23549;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)

DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 GN BGLC.
 OS Bacillus subtilis.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BSE616;
 RX MEDLINE: 9129280.
 RA Park S.H., Kim H.K., Pack M.Y.;
 RT "Characterization and structure of the cellulase gene of Bacillus
 subtilis BSE616.";
 RL Agric. Biol. Chem. 55:441-448(1991).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D01057; BAA00859.1; -
 CC PIR: J0111; J0111.
 CC HSSP: 006851; INBC.
 CC INTERPRO: IPR001547; -
 CC INTERPRO: IPR001956; -
 CC PRAM: PF00942; CBD_3; 1.
 CC PRAM: PF00150; cellulase; 1.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC K1 cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 499 ENDOGLUCANASE.
 FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 SO SEQUENCE 499 AA; 55169 MW; 2E821E3D8BAC04 CRC64;

Query Match 6.68; Score 171; DB 1; Length 499;
 Best Local Similarity 19.08; Pred. No. 0.00013;
 Matches 98; Conservative 104; Mismatches 189; Indels 126; Gaps 25;

QY 6 KIRISITLILLASSILF-----VSGTSTANANGS-FYVSGTTLTYDANGPEVWARG 55
 DB 2 KRISITITCLITLITLWGMGLASPAASAGKRTIPYAKNGQSLSTGQLYVNDGKAVQLKG 61
 QY 56 I-NHGHAMKYDQAT-TALEGIANTGANTV--RYLSDGGGWTKDDI-HTVRNITSLAED 109
 DB 62 ISSHGLQWYGEYVKNKDSIKWLRDMDGIVFRAAMTADGGIIDNPVKNKKKEAVEAKE 121
 QY 110 NMLVAVEVHATGYDSIASLNRADVYIEMRSALIGEDVYIINANEMG--SWED- 166
 DB 122 LGIVYIIDLWHLNDGNPNONEKKEFEKEM-SSLYGNTPVYIEIAEPENGVNWKMDI 180
 QY 167 -AMADGYKOAIPRLNAGLINTLWADAGWGFPSIDYREVENADPQ--RNTMESI 222
 DB 181 KPYAE---EYVSVIRKNDPNIIIV---GIGTWSQDVND-----AADQLKADNVMDAL 228
 QY 223 HMYEAGNASQVNTNIDVNLQDLALVIGEGHRTNGD---VDEA-TIMSYSEQGV 277
 DB 229 HF-YAGTHGQFLRKANYALSKGAPIVTEWGTSDASGNGGVFQDQREMLKYDSTKI 286
 QY 278 GWLAWSWKNGNPEWEYL---DLSNDWAGNNTNANGNTIVNGPYLRTSIR----- 324
 DB 287 SWVMNNLSDKOESSALKPGASKTGGMRLSDLSASGTVRENTILGTOKSTADIPETPAKD 346
 QY 325 -----LSTVFTGGSGDGTSPITLYDFEGSMQGTSSLSGGFWAVTEWSSKSHSLK 377

DB 347 KPTQENGISVOYRAG-----DGSN-----NSNOIR 371
 QY 378 ADIQLSNSQHYLVHVIQNTSLQ-----QNSRIQATVKHAMNGSVNGMTARLYV-- 426
 DB 372 PQQLKNNNG-----NTVYDLKDYARWYKANKKGAVDDCYAQLOGGNTYTFVYL 423
 QY 427 -KTGEGYTWY-----SGSFVPINGSGTTLSDLSN 456
 DB 424 HKPKQAGDYTLIEGFNGNTLAP--GASTGNIQRLHN 458

RESULT 15
 GUND_CLOCE
 ID GUND_CLOCE STANDARD; PRT; 584 AA.
 AC P25472;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE ENDOGLUCANASE D PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE D)
 DE (CELLULOSE D) (EGGCD).
 GN CELGCD.
 OS Clostridium cellulolyticum.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35319;
 RX MEDLINE: 92009193.
 RA Shima S., Igarashi Y., Kodama T.;
 RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
 RT celgcd, of Clostridium cellulolyticum.";
 RL Gene 104:33-38(1991).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIODHYDROLASES THAT CUT THE DISACCHARIDE CELLOBIOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE ST COMPONENT.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D90341; BAA14354.1; -
 CC PIR: J01229; J01229.
 CC HSSP: P17901; IEDG.
 CC INTERPRO: IPR001547; -
 CC INTERPRO: IPR002048; -
 CC INTERPRO: IPR002105; -
 CC PRAM: PF00404; Dockerin_1; 2.
 CC PRAM: PF00150; cellulase; 1.
 CC PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 CC PROSITE: PS00448; CLOS_CELLULOSEOME_RPT; 1.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC K1 cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 584 ENDOGLUCANASE D.
 FT DOMAIN 25 328 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 329 353 PRO/THR-RICH (LINKER).

FT DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 530 584 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 530 552 1.
FT REPEAT 562 584 2.
FT ACT_SITE 159 159 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 264 264 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 584 AA; 66061 MW; 0FC41257E81322C3 CRC64;

Query Match 6.5%; Score 170.5; DB 1; Length 584;
Best Local Similarity 20.5%; Pred. No. 0.00017;
Matches 119; Conservative 91; Mismatches 204; Indels 167; Gaps 30;

QY 6 KRIFFSTLLASSLIFVSGSTANANGFYVSGTLLYDA-NGNPFVNR---GINHG- 60
DB 2 KRIALLISCSTIMSPFMS-----VYGAINSDMKRMGIGMNLGNT 44
QY 61 -----ANYKQATTAIGIANTGANTVRIYVLSGGQWTKDDIHTVRNLSLAEDNHLV 113
DB 45 FDAPTEGSKAAQAEYFFDFKQAGFKHVRIP-----RW---DQHTLANSPTVDSNFTLN 97
QY 114 AVPEV-----HDATGYDIA-SLNRAVDYWIEMRSALIGKEDTVIINTA 156
DB 98 RIETVIDMELSRGFYVINSHDPTLMDNYSONIGRFEKIMEQIAQRFKRSENLYFEIL 157
QY 157 NEMFGMEGDADAGYKQAIPLRNAGLNHTLMDAAGWGQPPQSIHDYGREYFNADPQR 216
DB 158 NEPHGNTDSQINDMKRLINIRKTNPTRNVIIGAGYNSY-NSLSQI-EIPN-DP-- 211
QY 217 NTMESIHME-YAGNAGSOVRTNIDRVNLQDLALVIGFGRHRTNGDVDEATIMSYSEOR 275
DB 212 NLIATPHYDPSFTHQW-----GTWG---TKNDMD-ALAMAFNHYK 250
QY 276 GVGWLAWSKNGCPEN-----EYLDLSNDMA---GNMLTAMGNTIYNGP 316
DB 251 -----KMSDKNNIPVYLGEGYGVGHSDDRTSAVWPFDFVSDQALSHGFCGAMD---NGV 301
QY 317 YGL-----RETSR-----LSTVFTGGSDGSGTSPT-----TLX 344
DB 302 FGSVDNDMAFYNRDTRQDFKEILMALITTYGTYDWTPTPETNPDPPRTPATPAYGEOLIE 361
QY 345 DEFGSMQGTGSSLGCPWAVTEWSSKSGSHLKADITOLSSNOHYLHVIONTSLOON--- 401
DB 362 DEFGAMQ-W-AAISGVDAATASCCKISSGKSNNGLETITYAGSSNGTYGVNDENRNDWEK 418
QY 402 -SRIGATVHANWGSVYNGMTARLYYKGTWYSGSFVPINGSSGTTLSIDLSN--- 456
DB 419 WOKISFDIKSNTNEVRLIAEQSKIEGDEHM---TYIKPSTSMWTIEIPSSFTKR 475
QY 457 -----VQNLQVREI---GVQFQASDSSGQTSIYIDNV 487
DB 476 MDYQPPADGSETFDLKYKVSLLFMYSNNSG--TLINDNI 514

Search completed: December 19, 2000, 17:35:51
Job time: 976 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:42:13 ; Search time 248.55 Seconds

(without alignments)
125.108 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607

Sequence: 1 LNNFKRIPRITSLDASS.....QSASDSSGQTSIYIDNVIYE 490

Scoring table: BLOSUM62

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1614	61.9	516	2 JE0134	mannan endo-1,4-de
2	761.5	29.2	384	2 T37232	secreted beta-mann
3	731.5	28.1	363	2 S30386	mannan endo-1,4-de
4	638	24.5	1331	2 A48954	mannan endo-1,4-de
5	427	16.4	513	2 A37219	mannan endo-1,4-de
6	239.5	9.2	557	2 JCS487	cellulase (EC 3.2.
7	234	9.0	448	2 A25156	cellulase (EC 3.2.
8	226.5	8.7	409	2 B25156	cellulase (EC 3.2.
9	226	8.7	409	2 B25156	cellulase (EC 3.2.
10	220.5	8.5	505	2 S39962	endo-glucanase
11	207	7.9	504	2 S54744	cellulase (EC 3.2.
12	203.5	7.8	570	2 S56132	cellulase (EC 3.2.
13	193	7.4	499	2 A27198	cellulase (EC 3.2.
14	188.5	7.2	508	2 A26874	cellulase (EC 3.2.
15	181	6.9	508	2 G69593	cellulase (EC 3.2.
16	180	6.9	486	2 T40548	cellulase (EC 3.2.
17	179	6.9	466	2 C42360	cellulase (EC 3.2.
18	173	6.6	466	2 JCS467	cellulase (EC 3.2.
19	171	6.6	499	2 JN0111	cellulase (EC 3.2.
20	170.5	6.5	822	2 JN0129	cellulase (EC 3.2.
21	169	6.5	822	2 JN0611	cellulase (EC 3.2.
22	168	6.4	800	2 A28003	cellulase (EC 3.2.
23	166	6.4	941	2 S28043	cellulase (EC 3.2.
24	164.5	6.3	825	2 JCS0174	cellulase (EC 3.2.
25	160.5	6.2	510	2 T37541	probable glycolipid
26	156	6.0	611	2 JCS177	endo-glucanase V (E
27	155	5.9	357	1 C2CLEM	cellulase (EC 3.2.
28	150	5.8	357	1 PC4404	cellulase (EC 3.2.
29	147.5	5.7	429	2 S29044	endo-glucanase A pr

30	146.5	5.6	747	2	cellulase (EC 3.2.
31	140.5	5.4	426	2	cellulase (EC 3.2.
32	140.5	5.4	1335	2	glycoprotein VP260
33	135	5.2	438	2	glucan 1,3-beta-gl
34	133	5.1	1091	2	hypothetical prote
35	132.5	5.1	1608	2	hemolysin A - Ser
36	132	5.1	1300	2	probable serine pr
37	131	5.0	517	2	cellulase (EC 3.2.
38	130	5.0	515	2	endo-glucanase - C1
39	129	4.9	566	2	cellulase (EC 3.2.
40	129	4.9	1070	2	cellulase (EC 3.2.
41	129	4.9	1536	2	high-molecular-wet
42	129	4.9	2020	2	ABC-type transport
43	128.5	4.9	890	2	bacteriocin BCN5 -
44	127	4.9	32	2	guar gum-degrading
45	126.5	4.9	428	2	cellulase (EC 3.2.

ALIGNMENTS

RESULT 1
JE0134
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N/Alternate names: endo-1,4-beta-mannanase
C/Species: Bacillus circulans
C/Date: 03-Jun-1998 #sequence_revision 10-Jul-1998 #text-change 17-Mar-1999
C/Accession: JE0134
R/Yoshida, S.; Sako, Y.; Uchida, A.
Biosci. Biotechnol. Biochem. 62, 514-520, 1998
A/Title: Cloning, sequence analysis and expression in Escherichia coli of a gene cc
A/Reference number: JE0134; M01D:3623274
A/Accession: JE0134
A/Molecule type: mRNA
A/Residues: 1-516 <YOS>
A/Cross-references: DDBJ:AB007123
A/Note: the authors translated the codon CAA for residue 259 as Asn. CAA for residue
A/Note: the translation of the nucleotide sequence 1294-1365 is not given in this p
C/Keywords: glycoprotein; glycosidase; hydrolase

Query Match	61.9%	Score 1614;	DB 2;	Length 516;
Best Local Similarity	59.4%	Pred. No. 1e-102;		
Matches 291;	Conservative 84;	Mismatches 87;	Indels 28;	Gaps 3
QY	27	TSTANANGFYVSGTTLVDANGNPVFMGINHGMAWYDQATTAEGTANTGANTRVYL	85	
DB	29	TNKAAAGFYVSGTKLADATGQPPVFMGVNHAHWYDQASTALPAKAGANTRVYL	88	
QY	87	SDGGWTKDIDHTVRNLISLAEDNHLVAVPEYHDATGYDSTASLNRAVDYIEMRSALIG	146	
DB	89	ANGHKWTLDDVTVNNITLCEQNKLIVLEVHDATGSDSLDDNNAVYWGISAIG	148	
QY	147	KEDYIINIANWFGSWMGDMADGKQAIPLRANGLNHTLVNVAAGMGPDSIHXYG	206	
DB	149	KEDYIINIANWFGTWMGDMADGKQAIPLRANGLNHTLVNVAAGMGPDSVKNYK	208	
QY	207	REVNADPQNTMFSIHAYEYAGNASQRTNIDRVLNODLALVIGEGHRTNDVDEA	266	
DB	209	TEVNLADPLKNTVFISIHTEYAGNASVTKSIDVLLKNLALIIIGEGHRTNDVDEA	268	
QY	267	TIMSYSEORGVGLWMSKNGKNGPEWYLDLSNDNAGNNLWAGNTIVNGPYGLRSTSL	326	
DB	269	TIMSYSEORGVGLWMSKNGKNGSDSLAYLDMTDMAGNSLTSFGNTVNGSINKATSYLS	328	
QY	327	TYFTGGSDGSGTPT-----LTFEBSMGMGSSISG	360	
DB	329	GTF--GVTPTSSPTSTPTSTPTSPPTSPGNGNTLLDFEFTGNGSGNNISG	386	
QY	361	GPMATVENSCKSHLADIOSNSOYHLAVTQTSLOQSRIDQATVYKAMNSVNGM	420	
DB	387	GPMATVENSCKSHLADIOSNSOYHLAVTQTSLOQSRIDQATVYKAMNSVNGM	446	

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:07:33 ; Search time 362.48 Seconds
(without alignments)
126.223 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNCFKRFITSLSLASS.....OSASDSGGSITIDNYIVE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP invertebrate:*
- 6: SP mammal:*
- 7: SP mhc:*
- 8: SP organelle:*
- 9: SP phage:*
- 10: SP plant:*
- 11: SP rodent:*
- 12: SP virus:*
- 13: SP vertebrate:*
- 14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1622	62.2	516	2	066185 bacillus ci
2	814.5	31.2	395	2	069347 vibrio sp.
3	761.5	29.2	384	2	086599 streptomyce
4	634	24.3	279	2	092FI13 thermomonos
5	628	24.1	930	2	09RFX5 calidibacill
6	541	20.8	327	2	09RA15 streptomyce
7	253.5	9.7	1097	2	09RA17 anaerobic t
8	239.5	9.2	557	2	P94622 clostridium
9	228	8.7	389	2	059232 bacillus sp
10	203.5	7.8	570	2	059665 pseudomonas
11	193	7.4	499	5	045532 bacillus su
12	187.5	7.2	476	5	018453 heterodera
13	187.5	7.2	481	2	066064 actinomyces
14	187.5	7.2	501	2	083012 bacillus sp
15	186.5	7.2	387	2	031029 erwina car
16	185.5	7.1	930	2	059290 clostridium
17	180	6.9	486	2	045430 bacillus sp
18	180	6.9	635	2	066065 fibrobacter
19	179	6.9	783	2	045554 bacillus sp

20	178	6.8	749	2	059154 anaerocellu
21	177	6.8	499	2	052731 bacillus sp
22	173	6.6	621	2	007653 cellvibrio
23	169	6.5	821	2	059241 bacillus sp
24	161	6.2	478	5	016028 globodera r
25	160	6.1	278	2	09RKS5 streptomyce
26	152	5.8	506	5	09UA57 meloidogyne
27	148.5	5.7	470	5	09UM5 globodera t
28	145.5	5.6	395	5	09UM4 globodera t
29	140.5	5.4	1335	12	089353 paramacium
30	138.5	5.3	910	3	P87211 orphnomycs
31	136.5	5.2	494	2	086099 alteromonas
32	136	5.2	438	3	09UR18 candida alb
33	133.5	5.1	319	5	018454 heterodera
34	132	5.1	1300	2	032591 escherichia
35	132	5.1	1300	2	032555 escherichia
36	131.5	5.0	319	5	077449 heterodera
37	130.5	5.0	319	5	061595 heterodera
38	130	5.0	392	5	044078 globodera r
39	129	4.9	391	5	077094 globodera r
40	129	4.9	519	2	047916 fibrobacter
41	129	4.9	1070	2	055365 synchocyst
42	129	4.9	1107	2	P76088 escherichia
43	129	4.9	1167	2	P76859 escherichia
44	129	4.9	1536	2	048031 haemophilus
45	127.5	4.9	739	2	09X687 salmonella

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	516 AA.
066185	066185			
AC	066185			
DT	01-AUG-1998 (TREMBLrel. 07, Created)			
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DE	MANNANASE.			
OS	Bacillus circulans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
CC	Bacillus/staphylococcus group; Bacillus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 96233274.			
RA	Yoshida S., Sako Y., Uchida A.;			
RT	"Cloning, sequence analysis, and expression in Escherichia coli of a			
RT	gene coding for an enzyme from Bacillus circulans K-1 that degrades			
RT	guar gum."			
RL	Biosci. Biotechnol. Biochem. 62:514-520(1998).			
DR	EMBL: AB007123; BAA25878.1; -.			
DR	INTERPRO: IPR001547; -.			
DR	PFAM: PF00150; cellulase: 1.			
SQ	SEQUENCE 516 AA; 55245 MW; DBE5A48842AE8062 CRC64;			
Query Match 62.2%; Score 1622; DB 2; Length 516;				
Best local similarity 59.6%; Pred. No. 1.2e-107;				
Matches 292; Conservative 86; Mismatches 84; Indels 28; Gaps 4;				
QY	27	TSTAAANSFYSGTLLYANGNPFYMRGINGHANYKQATTAIGTANTGYRYLV	86	
DB	29	TNNAAASGFYSGTKLIDATGQPFYMRGVNNAHTYKQDLSTAIPIAKTGNTIRYLV	88	
QY	87	SDGQWTQKDIIVTRNLISLAEDNHLVAPEVDATGYDSIASINRAVDYIEMRSALIG	146	
DB	89	ANCHKRTLDVTVNNIITLCEONKLIATLVYHDAIGSSISLDNAAYVYIGIKSALIG	148	
QY	147	KEDTVIINIANEWGSGWEDANAAGYKQAIPLRLNAGLNLHTLVADAAGGQEPQSIHDYG	206	
DB	149	KEDRVIIINIANEWYGTWDGVAANAGYKQAIPLRLNAGLNLHTLVDSAGGQEPDSVKNYG	208	
QY	207	REVFNADPQARNMFSTIHMEIYAGNAGYRTNIDRYLNDLALVIGFGRHRTNGVDVDA	266	

```

Db      209  TEVLNADFLKNTVPSIHXYEYAGNASTVKSINIGVLTNKLNLALIGFEGGQHTNGVDYEA 268
QY      267  TIMSYSEBRGIGWILAMSKWKGKPMWEIIDLSDNDMAGNNLTMAGNTLYNGPYGLRETSRLS 326
Db      269  TIMSYSOQRKGWGLAMWSKWKGNSSDLAYLDMTNMAGNSLTISFGMTVYVNGSNGLKATSVLS 328
QY      327  TVTFEGSGSDGSGTPTT-----LYDFEGSMQMGWTSLSG 360
Db      329  GIF--GGVTFPSSPSTSTPTSTPTPTPTPTSPGNGTILYDFETGCGWGNNTISG 386
QY      361  GPMAVTEWSKSGSHSLKADLIQSSNSQHYLVHVIQNTSLQOONSRIQATVKHANMGVYNGM 420
Db      387  GPWVTNEMKATGAQDLKADVSLQSSNHSLSLTISNQMLSGSSSLAKVYKHAMNNGINIGI 446
QY      421  TARKLYVKTGHGYIYWSGFEVPFINGSGETTSLSDLSNVQNTLSQVREIGVQFQASADSSGQT 480
Db      447  YAKLYVKVSGMTWYDSEGNLIQSNDDGTLITLSIGTISNLSVKEIGVEFPRASSNSGQS 506
QY      481  SIYIDNVAYE 490
Db      507  AIVDVSVSLQ 516

```

Q	SEQUENCE	395 AA;	A3E30B3A978C5DA1	CRC64;
DR	PROSITE, PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.			
DR	PROSITE, 395 AA; 43098 NM; A3E30B3A978C5DA1			
DR	SEQUENCE			
OS	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; <i>Vibrio</i>			
OC	[1]			
GN	MANA.			
DE	BETA-1,4-MANNANASE.			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)			
DT	01-AUG-1998 (TREMBLrel. 07, Created)			
AC	069347;			
ID	069347	PRELIMINARY;	PRT;	395 AA.

Query Match	31.28;	Score 814.5;	DB 2;	Length 395;
Best Local Similarity	46.58;	Pred. No. 1.23e-50;		
Matches 168; Conservative	62;	Mismatches 104;	Indels 27;	Gaps 9

```

QY      6 KKFISFJLILLASSIFVSGSTPANNNGSGYGGTLLYDANGPFPMRGJNGHAYKD      65
      1 | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db      5 KAIRSL-----LLFIMASC---AHAGGYVNGVLYEANGSAFRIKIRINHAHYTD      52

QY      6 QATTAIGCIANTGANTYRIYLSDCGQNTKODIHVVRNLISLAEDNHLVAVPEVHADTGY-      124
      : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      53 KLSVALSGIATGANTYRYVLSNGYRMTKNDVSPVTIMILAKNNMLIALEVEDTGYG      112

QY      125 --DSIASLNRADVWIMRSLICKERTVIINLANEMFG--SWEGDAAADQYKOAIPLRN      161
      | | | : | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db      113 EESSAASLDSADQWILKKEMLIGELYIYLINLGNFEFGANNNDVAVANWDVSHIQLR      172

QY      182 AGLNHTLMDAGMGQ--FPOSIDHYGREVENADPQRTMTESIHYE--YAGGNSAQVETNT      239
      | | | | | | | | : | | | : | | | | : | | | | : | | | | : | | |
Db      173 AGINHITMVAPEMNGQWKGFMNLNAPFVNSDPKLTITISVHYEYVSYNS--VNDYI      230

QY      240 DRYVNLQDLAVLYGFGHRHNGVDVATITNSYSPQRBVGLAAMKNGCEWETYLSD      299
      | | | | | | | | : | | | : | | | | : | | | | : | | | | : | | |
Db      231 SSFTNNQLVLYIGFASHTKGADVDDESIHERSETTSLGIGTIGMSWSGNDTTTSDLDIVNN      290

```

```

QY 300 WAGNNLFAAGTIVNGPYGLEETRLSTVTTGGG---SPDGTSP---TLTLDGSGMGG 350
    | : : | | : : | : : | : | | | | | | | | | : : : |
Db 291 WDNNYSYTWGAVLLNGONGIKSTLTATVTCGACNDDBDSSGEPICSSSNAVDPDGDGWC 351
    | : : | | : : | : : | : | | | | | | | | | : : : |
QY 353 W 353
    | : : | | : : | : : | : | | | | | | | | | : : : |
Db 351 W 351
    | : : | | : : | : : | : | | | | | | | | | : : : |

```

RESULT	3		
086599			
ID	086599	PRELIMINARY;	PRT; 384 AA.
AC	086599;		
DT	01-NOV-1998	(TREMBLrel. 08, Created)	
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)	
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)	
DE	SECRETED BETA-MANNOSIDASE.		
GN	MAN2.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinomycetidae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RA	Oliver K., Harris D.;		
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RA	Parkhill J., Batteil B.G., Rajandream M.A.;		
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.		

RC STRAIN-A3(2);
RA MEDLINE: 97000351.
RA Kienast H., Kleiser H.M., Denapatre D., Eichner A., Cullum J.,
RA Kinast H., Hopwood D.A.;
RT "A set of ordered contigs and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A5(2) chromosome";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL003154; CAA20610.1; -.
DR INTERPRO: IPR001547; -.
DR INTERPRO: IPR002883; -.
DR PRAM: PFO0150; cellulase; 1.
DR PRAM: PFO2013; CBD_5; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
Q0 SEQUENCE 384 AA; 39847 MW; B8AD156E3FD3BD11 CRC64;

Query Match	29.28;	Score 761.5;	DB 2;	Length 384;
Best Local Similarity	44.28;	Pred. No. 1.3e-46;		
Matches 161;	Conservative 59;	Mismatches 121;	Indels 23;	Gaps 8;

[illegible]

09RJ15
ID 09RJ15 PRELIMINARY; PRT; 327 AA.
AC 09RJ15
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUTATIVE SECRETED BETA-MANNOSIDASE (FRAGMENT).
GN MANA.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kleser H.M., Denepalte D., Eichner A., Cullum J.,
RA Kinsahl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL121746; CMB57406.1; -.
DR INTERPRO; IPR001547; -.
DR PFAM; PF00150; cellulase; 1.
FT NON_TER 327
SQ SEQUENCE 327 AA; 34772 MW; A21B0068C2C0975D CRC64;

Query Match 20.8%; Score 541; DB 2; Length 327;
Best Local Similarity 41.5%; Pred. No. 5e-31;
Matches 117; Conservative 54; Mismatches 101; Indels 10; Gaps 8;

QY 25 SGTSTANAN-SGPFYSGTLLDANGNPFYMRGINHGHAMTKQATTAIGTANTGATYR 83
DB 47 AGPSSGAGAAAGLHIGDGRLENGNDFYMRGVNHAHTYPER-TQSLADYALGANSVR 105
QY 84 IYLSDGQGTCKDITHTYRNLISLAEDNHLVAPEVDATGY--DSIA-SLNRAVDYWM 140
DB 106 VYLSGSHRSENGPADVAIVIEGCKANRLICYLEVDTTGYADDAAGTLDAADYIGL 165
QY 141 RSALIGKEDTVIINANEFWSGEGDAMADYKQALPRLNAGLNTLWYDAAGWGQFPQ 200
DB 166 KQVLAGEQEDYVIVNIGNEPFGMTDPAGWTEPTVAAVKTLRAAGLQHTINVDAPNMQDQW 225
QY 201 SI-HDYGREVFANADPORNMFSTHMEYAGNASOVRTNIDVNLNDLALVIGEF-GHH 258
DB 226 GYMRANANRKYVADDPGNLIFSTHMS-VFTRAQETTYDLNLFVADLELPILIGFEGPAD 284
QY 259 TNGDVEATIMSYSEQGYGLWMSKNGGPEMEYDLSDNW 300
DB 285 QYGDPEDETMATAEQLRGLYGLWMSGNTD--PYDLALDF 324
RESULT 7
Q9ZAI7 PRELIMINARY; PRT; 1097 AA.
AC Q9ZAI7
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE S-LAYER ASSOCIATED MULTIDOMAIN ENDOGLUCANASE.
GN CELA.
OS anaerobic thermophile KM-THCJ.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.
RA [1]

RP SEQUENCE FROM N.A.
RC STRAIN-KM-THCJ;
RX MEDLINE; 99175466.
RA Camm I.K., Kocherginskaya S., King M.R., White B.A., Mackie R.I.;
RT "Molecular cloning, sequencing, and expression of a novel multidomain
RT mannanase gene from Thermomonas bacterium polysaccharolyticum.";
RL J. Bacteriol. 181:1643-1651(1999).
DR EMBL; 082255; AAD09354.1; -.
DR INTERPRO; IPR001119; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002088; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00395; SH; 3.
DR PFAM; PF02018; CBD_6; 2.
DR PROSITE; PS00904; PPTA; UNKNOWN_1.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
SQ SEQUENCE 1097 AA; 119762 MW; C41CEB92C0494B9 CRC64;

Query Match 9.7%; Score 253.5; DB 2; Length 1097;
Best Local Similarity 22.3%; Pred. No. 7.8e-10;
Matches 104; Conservative 76; Mismatches 181; Indels 105; Gaps 20;

QY 8 IFSITSLIASLIFVAGSTANANSG--FYVSGTLYDANGNPFYMRGIN-HGH-AW 62
DB 10 VFVMTLYALNGVFIYGKQAKAGTSGDGRFHYGKNIVDPDGNDFYKGVNIGYRISW 69
QY 63 YKQATTAIEGANTGA-NYVRIYISDG-----GQWTKDDIHTYRNLISLAEDNHLVAV 115
DB 70 EKRSTYADVHLIADYWKENTVRLNCFIQNNWMBEGTGANNIDAI--IKAFKAKYVE 126
QY 116 PEVHATGYDSA-----SLNRAVDYWMKSAIGEDYITIANEMGSGWE- 164
DB 127 IDLHDTTGIPLSNPPRPGOSLDOAI-AMFKELAAATYKONPYWENTNPESSITAPL 185
QY 165 GDAMADYKQALPRLNAGLNTLWYDAAGWGQFOSIHD-----YGREVF 210
DB 186 DPQGVANEEIITKATRSRGADNIIYVD--GWSYANEGIEQNTPTVDEKRSALVITYGDL 243
QY 211 NADPORNMFSTHMEYAGNASOVRTNIDVNLNDLALVIGEGHRTNGVDDEATIMS 270
DB 244 NADSAKNTIIFEFHNYN-BGDIQKVEDYIDRANKGLYVFMEYEGKDS--DAAKEGVKS 300
QY 271 YSE---ORGVGLWMSKNGGPEMEYDLSDNWAGNNLTANGNTIYVPGYLRSTRST 327
DB 301 GLQAVMKAGAGRIYNNMG---YDLDTLSG----- 328
QY 328 VFTGGG-----DGTSPITLYDFEGSMQGTSSISGCPWAVTESSKSGSHLIKADIQ 381
DB 329 --TGRSGMEIKTKDGSKPTNL-----SWVGDKIMDDNHGIT--PTFDDQPKVDLA 376
QY 382 LSSNQHTLHYTQNTSLQNSRIQATYKHAHWGSGVGNMTRKLYK 427
DB 377 LER-----LIANNGFAGDKVOFTFLRNSGDLPIGDKSKVYVK 416
RESULT 8
P94622
AC P94622 PRELIMINARY; PRT; 557 AA.
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ENDO-1,4-BETA-GLUCANASE ENGf (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
GN ENGf.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97136706.
RA Shewetta S.A., Ichi-ishi A., Park J.S., Liu C., Malburg L.M.,

Doi R.H.:
 RA "Characterization of engf, a gene for a non-cellulosomal Clostridium
 RT cellulovorans endoglucanase.";
 RL Gene 182:163-167(1996).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: U37056; AAB40891.1; -.
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; -.
 DR PFM: PF00150; cellulase, 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Hydrolyase; Glycosidase.
 KW SEQUENCE 557 AA; 60131 MW; D186EC88E504EED CRC64;
 SO

Query Match 9.2%; Score 239.5; DB 2; Length 557;
 Best Local Similarity 23.1%; Pred. No. 3e-09;
 Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

DB 2 NNGFKKTSI-----TLSLILASSILFVSGTS-----TANAN-----SGFTV 38
 DB 3 NNVKKKILSIYAAGAMLMALPNNVVAETTSNLTGNANVKKPSVGRKLQLLNKNGI-- 60
 OY 39 SGTLYDANGNPFVARGIN-HGHANYKQDATIAGIANTGA-----NTVRIYL- 86
 DB 61 --KILCDKQGNFIQLRGSTHGLQNF-----PGVANNNAFALISNDWNSNVIKLAAY 110
 OY 87 -SDGQWMTKDI-HTVNRLISLAEDNHLVAPEVDATGYDSIASINRAVDWIERSAL 144
 DB 111 VAEGVATNPVKQVINGINAIANDMVIYDWMHMGNDNVAISGASQEFNDISLT 170
 OY 145 IGKERTVLIINANEFGSNEG---DA--WADGYKQAIIP---RLRNAGLNLHLMVDAAGW 196
 DB 171 YPNKNRIIYECLNEPENGNGVTNDATGMAQYKSYATPIVQLLRKNGENLIIYGNPFW 230
 OY 197 QEPQSHDYGREVFADPQBMTEFISIMYEYAGN-ASQVETNIDR-----VINQDA 248
 DB 231 QRPDLAAD-----NPIDSNMTSYHF--YSGINPISIVTDNRDANSNVRALHGA 282
 OY 249 LVIGEGFRHTNGVDYDEATIMSYSBOGVWLAMSKNGPMEYLDLSND----- 299
 DB 283 VFATEMGTSLATG-----TTGPYL-AKADAWLDF-LNGNNISMCNPSISNKEKAALNS 335
 OY 300 -----WAGNNLTMAG-----NTIVGPGYGLRSTSLSTVFTGG 332
 DB 336 LTSIDPGSKLMADNELTSGQYRARIKGAYATPVDPVTQPPAPKDFSGGFDFENG 395
 OY 333 GSDG-----GTSPITLYDFEGSMQGTSSLSGPAVTEMSSKGSLSK-----ADIOS 383
 DB 396 TTQGFQVPPDSFITAIVENANNALKISNL-----NKGSDNLSSEGNFMANVRIS 445
 OY 384 SNS-OHYLVAVIONTSLOANSRIQATVKRANMGVNGMTARLYVTGHOY-----TW 434
 DB 446 ADIMGOSINITYGDTKLTMVLAIPV-----NVSIAAIPOSSTHGMGPTAIRW 496
 OY 435 YSGSTVPINGSSGTTLSLDSLVQNLISOYREIGVQFQASDS-----SGTSTIYI 484
 DB 497 TNNFVAQTDGYKATLTISTDSNPNFTIA-----TDAASVVTNMLIFVSGSDNISTL 550
 OY 485 DNV 487
 DB 551 DNT 553

RESULT 9
 AC 059232 PRELIMINARY; PRT: 389 AA.
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE ENDO-BETA-1,4-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULASE)
 DE (ENDOGLUCANASE) (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE).

OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-186-1;
 RA Sanchez Torres J;
 RL Theiss (1994), Universidad de Salamanca, Spain.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: 23876; CA83942.1; -.
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; -.
 DR PFM: PF00150; cellulase, 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Signal; Hydrolase; Glycosidase.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 ENDO-1,4-BETA-GLUCANASE.
 SQ SEQUENCE 389 AA; 43688 MW; 9111397485609B5F CRC64;
 SO

Query Match 8.7%; Score 228; DB 2; Length 389;
 Best Local Similarity 23.6%; Pred. No. 1.2e-08;
 Matches 100; Conservative 65; Mismatches 154; Indels 104; Gaps 21;

OY 6 KRIEITSLIASSILFVSGTSTANANS-----GTVSGTLYDANGNPFVNRGI-NH 58
 DB 2 KRITTFIVLMTVALFSIGNTTADNDVSVEHQSLSISNGELVNERGEQVQLKMSH 61
 OY 59 GHAWKQDATIIEGIANTGATVRIYLSDGQWTKD-----IHT----- 99
 DB 62 GLQWVG-----QFVNESKMLRDWGITVFPRAKITSSGGYIDPS 103
 OY 100 ---YRNLSLAEDNHLVAPEVDATGYDSIASINRAVDWIERSALIGKEDYIINI 155
 DB 104 VAEKTKAEVAALDDIYIIDMHLISDNDPRITYEKAQFDEH-SELGYDPVITYI 162
 OY 156 ANEWFGSWEGDAMADGYR---QAIPLRNAGLNLHLMVDAAGWQFPOSIDYGREVF 211
 DB 163 ANEPNPS--DYVMDRIKRYAEVIVIRKNDPNIIIV---GTGTSQDVH-HAADNL 216
 OY 212 ADPQNTFESIMYEYAGNSAQVETNIDRVNODLAVIGEGFRHTNGD-----YDEA 266
 DB 217 ADP--NVMIAFHF--YAGHGNLQNDQVDYALDQGAALFVSWGTSATGDCGVFLDEAQ 272
 OY 267 TMSYSEQGVWLAMS-----WKGNGPMEYLDLSNDWAGNNLTMAGNTIYNG 315
 DB 273 EWIDMDERNLSMAWSLHKESSAALMPGANP-----TGWTEALSPSGTFVRE- 324
 OY 316 PYGLRETSRLSTVFTGGSDGT---SPTIYDFE-----GSM---QGTGSSLSG--- 360
 DB 325 --KIESASIPSDPTPSPDPGEYPAWPTQIYTNELIYVHNGQLQWAKWTONQPGDXY 382
 OY 361 GPW 363
 DB 383 GPW 385

RESULT 10
 AC 059665 PRELIMINARY; PRT: 570 AA.
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE) (ENDOGLUCANASE)
 DE (CARBOXYMETHYL CELLULASE).
 GN CELL.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 CC Pseudomonas.
 RN [1]
 DE SEQUENCE FROM N.A.

RX MEDLINE: 95366948.
 RA Hall J., Black G.W., Ferreira L.M.A., Millward-Sadler S.J.,
 RA All B.R.S., Hatlewood G.P., Gilbert H.J.,
 RT "The non-catalytic cellulose-binding domain of a novel cellulase from
 RT Pseudomonas fluorescens subsp. cellulosa is important for the
 RT efficient hydrolysis of Avicel."
 RL Biochem. J. 309:749-756(1995).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: X86798; CA60493.1; -
 DR HSSP: P07103; 1EG.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001919; -
 DR INTERPRO: IPR002883; -
 DR PFAM: PF00150; cellulase; 1.
 DR PFAM: PF00553; CBD_2; 1.
 DR PFAM: PF02013; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Hydrolase; glycosidase.
 KW SEQUENCE 570 AA; 59925 MW; D00B26F13D57BE7 CRC64;
 SQ

Query Match 7.8%; Score 203.5; DB 2; Length 570;
 Best Local Similarity 23.3%; Pred. No. 1.1e-06;
 Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

QY 14 SLLASSILFVSGSTANANGSGFYVSGTLLDANGNPFVARGIN-----HGHWYKD 65
 DB 22 AIIILGSGELIGVSAQADYAPLSPVGNKIL-RANGOPASFGSGLSWSTWEGGEYVNA 80
 QY 66 QATTALEGIANTGANTVRIVL---SDGGQWT---KDIHTVNNLSLAEDNHLVAVPEY 118
 DB 81 QVVSMLK--SDWAKLVRAANGVEDEGGTLDPANDRYTOYVD-AALINDMVIIDMS 137
 QY 119 HCATGDSIASLRANDYIEKRSALIGEDYIINIANEMFG-SWEGDAMADGIKQA-I 176
 DB 138 HNAHQQS-----QALAFQEM-ARKYGANNHYIEIYNEPLQVSWNT--IKPYAQAVI 189
 QY 177 PRLRNAGLHNTLMVDAAGGQFPQSIHDYGREYFNADP---QRNTFSIHMEYVAGNAS 233
 DB 190 AALRAIDPNLILVGNPTMSQ-----DQVYANDPITGTONIAYTLHF--YAGTHGQ 239
 QY 234 QVRTNIDRYLNODLALVIGFGRHTNGVDENATIMSYSEORGVGLWAMSKNGPEWEY 293
 DB 240 YLRDAQRLNGLALFVEMGSVYANGD---GAVANSF----- 275
 QY 294 LRLSDMAGNNLTMANGNTIVNGPYGLRETSRLSTVFTGGSGDGSPTLLYDFEGSMQGW 353
 DB 276 ---TNAWVSEKMT---NHTISNANMALNDKVEGASALVPGASANG-----GW 315
 QY 354 TGSSSL-SGGPMA---VTEW---SSKGSLSLRADIOSNSQHYLHVIONTSILOQNSRIQ 405
 DB 316 VNSQGLASALAKSLISGMPSYNTSSSSSAVSSQOVSSSQ--APVSSSSSTIASVVS 373
 QY 406 ATV--KHANWGVGNGMTARLYVKTGHGYT-----YSGSPVPINGSSTLTSL 452
 DB 374 SAVSGQCCNMW---YGLTYPLCSTTNGWEMNASCIRATCSQGPAPWGIWGSTSS- 428
 QY 453 DLSNVONTSOVEIGVQPOSASDSSGQTSI 482
 DB 429 ---QASSVYSSSSSLVSSSSSSSV 453

RESULT 11
 ID 045532 PRELIMINARY; PRT: 499 AA.
 AC 045532.
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE CELLULASE.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RX MEDLINE: 87190397.
 RA Nakamura A., Uozumi T., Teruhiko B.,
 RT "Nucleotide sequence of a cellulase gene of Bacillus subtilis."
 RL Eur. J. Biochem. 164:317-320(1987).
 DR EMBL: M28332; AA22307.1; -
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001956; -
 DR PFAM: PF00150; cellulase; 1.
 DR PFAM: PF00942; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM: PD001947; -1.
 SQ SEQUENCE 499 AA; 55075 MW; D424AB9E63B94D23 CRC64;

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 5.2e-06;
 Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

QY 6 KRIFETLSLASSILF-----VSGSTANANGS-FYVSGTLLYDANGNPFVARG 55
 DB 2 KRISIFITCLILITVLIMGGLASPAASAGKTRPAKKGSLIKGTOLVYNDGRAVOLKG 61
 QY 56 I-NHGHWYKQDATTALIGIANTGANTVRIYVSDGGQTKD-----IHT----- 99
 DB 62 ISSHGLQWYGD-----FVNKSLIKWLRLDWDGITYPRAMTYADGGYI 103
 QY 100 -----VRNLISLAEDNHLVAVPEVDATGYSIASLNAVYVIMBRSALLGKEPTY 151
 DB 104 DNPSTVKNKVEAWEAKELGIYIIDWHILNDGNPNQHKERAKDFEKM--SSLVGNPNV 162
 QY 152 IINIANEMFG--SWEGD--ANADGYKQALPRLRNAGLHNTLMVDAAGGQFPQSIHDYGR 207
 DB 153 IYEINENPNGVNWKMRDIPYAE--EIVSIRKNDPNILITV---GIGTMSQVND--- 213
 QY 208 EYFNADPQ---RNTMFSIHMEYVAGNASQVRTNIDRYLNODLALVIGFGRHTNGD-- 262
 DB 214 ---AADQDKDANWYALHF--YAGTHGQSIRDRANVALSKGALFVTEMGTSDASGNG 268
 QY 263 ---VDEATIMSYSEORGVGLWAMSKNGPEW-EYLDLSN-DMAGNNLTMANGNTIVNGPYG 318
 DB 269 VFLDQSR-----EWLNTLDSKNISWVWNL----- 293
 QY 319 LRETSRLSTVFTGGSDGSPPTLLYDFEGSMQGTSSLSGGPWAYTESSKGSLSIKA 378
 DB 294 -----SDRQESSALK-----PGASKTGG-WPLLDLTLASGTF-VRE 327
 QY 379 DI--QLSNSQHYLHVIONTSILOQNSRIQATVKHANWGVGNGMTARLYVKTGHGYT--- 433
 DB 328 NILGNKSTKRPETPADNDPAENG-ISVQYKAGDGVNSQIRPDLHKNNGNATVLD 386
 QY 434 -----WYSGSPVPINGSSTLTSLDSNY 457
 DB 387 KDVTARATYNAK-----NKGQNPDCDYAQI 411

RESULT 12
 ID 018453
 AC 018453 PRELIMINARY; PRT: 476 AA.
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE BETA-1,4-ENDOGLUCANASE-1 PRECURSOR (EC 3.2.1.4) (CELLULASE)
 DE (EMDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE).
 GN HG-ENG-1.
 OS Heterodera glycines.
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 98226740.
 RA Smant G., Stockermans J.P.W.G., Yan Y., De Boer J.M., Baum T.J.,
 RA Wang X., Hussey R.S., Gommers F.J., Henriksat B., Davis E.L.,
 RA Helder J., Schots A., Bakker J.,
 RT "Endogenous cellulases in animals: Isolation of beta-1, 4-
 RT endoglucanase genes from two species of plant-parasitic cyst
 RT nematodes";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4906-4911(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-OP50;
 RA Yan Y., Smant G., Stockermans J.P.W.G., Wang X., Hussey R.S.,
 RA Bakker J., Helder J., Schots A., Davis E.L.,
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: AF006052; AAC48327.1; -
 DR EMBL: AF052733; AAC15707.1; -
 DR HSSP: P07103; 1EG2.
 DR INTERPRO: IPR001347; -
 DR PFAM: PF00150; cellulase; 1.
 DR SIGNAL: Hydrolase; glycosidase.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 476 BETA-1,4-ENDOLYCANASE-1.
 SQ SEQUENCE 476 AA; 49815 MW; 2056F90CA0783FE7 CRC64;

Query Match 7.2%; Score 187.5; DB 5; Length 476;
 Best Local Similarity 23.7%; Pred. No. 1.2e-05;
 Matches 120; Conservative 61; Mismatches 211; Indels 115; Gaps 23;

QY 13 LSLILASSLITFSG-TSTANANSFGVSGTLLDANGNPFVNGIN-HGHATKD-QAT 68
 DB 13 LFLLLALCALVSLTAVALPFGQLSVSGTNLGVANGOPVQLIGNSLFHWQVPOFWMND 72
 QY 69 TALEGIANTGANTVR--IYLSDGQWTKDDIHTVRL--ISLARDNHLVAVPEVDAT 122
 DB 73 TYKALKCNANNAVIRKANGVDEGGYLS--DANTAYILMAVAIVIAISNGITYLVDMHAN 130
 QY 123 GYDSIASLNRADVWYTEMSSALIGKEDYIINIANEMFG-SWEGDAMADGKQAIPLRN 181
 DB 131 SHP-----DEAVFFTRIQA-YGSRPHLYEDFNEPLSVSWT-DVLVPHKKVIAIRA 183
 QY 182 AGLNHTIAMDAAGWQ-----FQSIHDYGRFNADPQANTFSLHMETYAGNASSQVR 236
 DB 184 IDKKNVILIGTPTWSDVDVAQNPIKDY-----QNLMTYTLHFY-----ASSHP 227
 QY 237 TN-----IDRVLNODLALYIGERG-HRHTNGDVEDATIMSYSEQGVGW-LAMSWKNG 288
 DB 228 TNDLGAKLKTAVANGLPVYETETGCEASGNGNLNDSMSW-----WTLDSLKIST 280
 QY 289 PEMEYLDLS-----NDMAGNNLTAMGNTIVNGPYGLRSTSLSTVFTGGSD 335
 DB 281 ANMAISDFSEACALSPGTTAAVGVSSRWTSNGNVAAS--YKKKSTGVSCGSSSSGS 338
 QY 336 GGTSPITLIDFESMGWGTGSSLSGCPWATEWSKGSKLADQLSSNGHTLVAYION 395
 DB 339 SSSSSSSSSSSSSSSSSSSSSG-----SSSSSSSSSSSSSSSSSSSASISVYPSN 390
 QY 396 TSLQNSRIQATYKHAMGVSNGMTARLYKTHGYTWYSGSFYINGSSGTTLSLDS 455
 DB 391 T---WNGGQVNFETKNTISVP-----LCGVVFSVSLPSGTTL----- 425
 QY 456 NVQNTLSQVREIGVQFQASDSSGQTSI 482
 DB 426 -----GGSNMESASGQYSL 441

RESULT 13
 ID 066064 PRELIMINARY; PRT; 481 AA.
 AC 066064;

DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE ENDOLYCANASE.
 OS Actinomyces sp. 40.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-40;
 RA Cho K.K.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U94825; AAC06196.1; -
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001347; -
 DR INTERPRO: IPR001919; -
 DR PFAM: PF00553; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL HYDROL. F5; UNKNOWN 1.
 SQ SEQUENCE 481 AA; 50750 MW; 50A6A29BEE44057A CRC64;

Query Match 7.2%; Score 187.5; DB 2; Length 481;
 Best Local Similarity 20.6%; Pred. No. 1.2e-05;
 Matches 100; Conservative 76; Mismatches 212; Indels 97; Gaps 21;

QY 38 VSGTLLYDANGNPFVNGIN-HGHANYKDQAT-TALEGIANT-GANTVRIVL-----S 87
 DB 45 LSAPTILIDHEGMFQLRGASTHGLQWFGQYVKNKARQSLRDEGLMVRILAVYPREGYL 104
 QY 88 DGGQWTKDDIHTVRLNISLAEDNHLVAVPEVH-----DANGYDSIASLNRADVWY 139
 DB 105 QGSQAQMD--KTLGGVQATATDGMATYIIDWHVLYNTPNGDATQAQS-----FEK 152
 QY 140 MRSALIGKEDYIINIANEMFGSWEGDAMADGKQAI-----PRLNAGLNHTIAMD 191
 DB 153 SYAAKKSQVNGVIFEVCHNEPTGT-----PYDSGNDIYCYTRMAKAIADAGSDAILIG 208
 QY 192 AAGWQGFQSIHDYGRFVNADPQNTFSLHMETYAGNASSQVRTNIDRVLNODLALVI 251
 DB 209 TWTWS---QDIPAVACKPISADGFDMIVLHF--YAAHKKDLRAKLTALNAGTPVY 263
 QY 252 GERGF-HRHTNGDVEDATL--MSYSEQGVGLAMSWKNGPEWY-----LDLSNDWAG 302
 DB 264 SEFGLDASGNGIDODSNAAMWTLIAHNNISTAAALSNKRETAFFRPSVATISKWTG 323
 QY 303 NMLTAMGNTIVNGPYGLRSTSLSTVFTGGSDGTSPTLLYDFESSMGWGTG--SSLG 360
 DB 324 DDLTPSAIWLVTNRSRLADFDHAASGTSSGSKASSGTGAGNGKSSASASSPFAAG 383
 QY 361 GPWAV---TEWSSKGSLSKLADQLSSNSOHLVAYIONTSIQQNSRIQATYKHAMGVS 416
 DB 384 GLATATAIRNQNMSGATYAI--TVSNASGSKH-----EGAMQVTPDLDAEADI-WG-- 432
 QY 417 GNMTRARLYKTHGYTWYSGSFYINGSSGTTLSLDSLVNQVLSVYRITGVQFQASDS 476
 DB 433 -----GTVSHGTHYVAVAPDMNTALEGASAEIGFN---ASS 468
 QY 477 SGQTS 481
 DB 469 TQDAS 473

RESULT 14
 ID 083012 PRELIMINARY; PRT; 501 AA.
 AC 083012;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE CELULOSE.
 OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-5H;
 RA Khanongnuch C., Ooi T., Kinoshita S.;
 RT "Cloning and nucleotide sequence of beta-mannanase and cellulase gene
 from Bacillus sp. 5H."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016164; BAA31712.1;
 DR HSSP: 085465; 2A3H.
 DR INTERPRO: IPR001547;
 DR INTERPRO: IPR001956;
 DR PFAM: PF00942; cellulase; 1.
 DR PFAM: PF00942; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM: PD001947; -; 1.
 SQ SEQUENCE 501 AA; 55422 MW; F76D260A901E2D1D CRC64;

Query Match 7.2%; Score 187.5; DB 2; Length 501;
 Best Local Similarity 21.4%; Pred. No. 1.3e-05;
 Matches 86; Conservative 72; Mismatches 140; Indels 103; Gaps 19;

QY 6 KRTFTSLILASSILFVSGTSTANANSGFVSGTLYDANGNPFVNR 54
 DB 2 KRVSTILYVCLLI-SILATSGLGNPVASATGATPAVINGQLSINKTOLINQKAVOLK 60
 QY 55 GI-NHSHAWYKQDQATTAIEGANTGANTVRIYVLSGGQWTKD-----IHT----- 99
 DB 61 GISSHLOWGQD-----FVAKDSIKMLRDWGLITVFAAMYTADGGY 102
 QY 100 -----VRNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWIEMSKALIGKEDT 150
 DB 103 IDNPSRKNKYEAVEAKEIGIYIIDWHILNNGNPNQNEKAKEFEKETS-LYGNTPN 161
 QY 151 VIININAMWFG--SWBGD--AWADGYKQAIPLRLNAGLNHTLMDVADAGMGQFQSIHDYG 206
 DB 162 VITEIEEPNGDVNWKRDIPYAE--EVIATYRKNDPNTIIV--GISTWQSDVND-- 213
 QY 207 REVENADPQ---RNTMESIMTEYAGGNASQVRIINIDRVLNODLALVIGEGHRTNGD- 262
 DB 214 ---AADDOUKDANVYALHF--YAGTHGOSLRDKANYALSKGAPFVETMGSDASG 267
 QY 263 ---VDEA-TIMSISQDGVGWLAMSKNGPENEYL---DLSDNAGNLLTANGTIYN 314
 DB 268 GVFLDOSREWLNTLDSKRIISVMWNLSDKQESSALKPGASKTGWPLTDLTASGTFVRE 327
 QY 315 GPYGLRETSR-----LSTVFTGGSDGSGTS 339
 DB 328 NITGNKSTKERPETPAQDNPAQENGISVQKAG--DGIVN 366

RESULT 15
 ID 031029 PRELIMINARY; PRT; 387 AA.
 AC 031029;
 DT 01-JAN-1998 (T-EMBLrel. 05, Created)
 DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last annotation update)
 DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
 GN CELA.
 OS Erwilia carotovora subsp. carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwilia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LY34;
 RA Park Y.W., Lim S.T., Yun H.D.;
 RL Mol. Cells 0:0-0(1997).
 DR EMBL: AF025768; AAC02964.1; -;
 DR HSSP: 085465; 2A3H.
 DR INTERPRO: IPR001547; -;
 DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Signal; Hydrolase.
 FT SIGNAL 1; POTENTIAL.
 FT CHAIN 32 387 BETA(1,4)-GLUCAN GLUCANOHYDROLASE.
 SQ SEQUENCE 387 AA; 42003 MW; 6302B3DF8AC2738B CRC64;

Query Match 7.2%; Score 186.5; DB 2; Length 387;
 Best Local Similarity 21.4%; Pred. No. 1.1e-05;
 Matches 82; Conservative 63; Mismatches 143; Indels 95; Gaps 16;

QY 6 KRTFTSLILASSILFVSGTSTANANSGFVSGTLYDANGNPFVNRGI-NHSHAWY 63
 DB 10 KRLTGAVTTVIGMSLSTALSATVEYTHGQISNGRLVDEQGRVYLRGVSSHLOWF 69
 QY 64 KQATTAIEGIANGTANTVRIYVLSGGQWTKD-----IHTV 100
 DB 70 GD-----YKKDSMKMLPDDWGINVSVAMYTADGYISKPLANKV 111
 QY 101 RNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWIEMSKALIGKEDTYININAMWF 160
 DB 112 KEAVVAAOSIGYIYIIDWHILSDNEPNYKKEQATFEAM-AGLYGNSPTYIYETIANEPN 170
 QY 161 G--SWBGD--AWADGYKQAIPLRLNAGLNHTLMDVADAGMGQFQSIHDYGREVENADPQR 216
 DB 171 GGYTWDELRPFA---LEVETIRSKDPDNLIV--GISTWQSDIHD--AADNQLPDP 221
 QY 217 NTFMESIMTEYAGGNASQVRIINIDRVLNODLALVIGEGHRTNGD-----VDEATIMSY 271
 DB 222 NTLVYALHF--YAGTHGOFIRIDYAOISGALIFVSEWGTSDASONGGPFLEPESHTWDF 279
 QY 272 SPQRGVGWLAMSKNGPENEYLDLS-----NDWAGNLLTANGTIYNPGYG 318
 DB 280 LNNRG-----GSRVNSILTDKSEASALAPGASKCGWTEDNLSASGKVF----- 324
 QY 319 LRETSRLSTVFTGGSDGSGTSPT 341
 DB 325 -RAQIRAAATLSGC--DPTTTP 344

Search completed: December 19, 2000, 17:32:55
 Job time: 1522 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:00:14 ; Search time 226.76 Seconds
(without alignments)
73.889 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607

Sequence: 1 LNNNGFKKIRTSITSLILASS.....QSAQSDSGGTSTIYDNYIV 490

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36:*

1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2607	100.0	490	Y54122	Amino acid sequenc
2	1804	69.2	493	Y44495	Bacillus agaradher
3	1804	69.2	493	Y54124	Amino acid sequenc
4	1711	65.6	468	Y44496	Bacillus agaradher
5	1711	65.6	468	Y54125	Amino acid sequenc
6	1640	62.9	476	Y54123	Amino acid sequenc
7	1152	44.2	331	Y54127	A mannase-linker
8	1075	41.2	369	Y54129	Amino acid sequenc
9	957.5	36.7	320	Y54132	Amino acid sequenc
10	750.5	28.8	305	Y54130	Amino acid sequenc
11	546	20.9	188	Y54133	Amino acid sequenc
12	419	16.1	487	P91007	Beta-mannase of al

13	318	12.2	132	21	Y54131
14	229.5	8.8	400	18	W23601
15	228.5	8.8	400	18	W22521
16	228.5	8.8	400	19	W57431
17	229.5	8.8	462	19	W57433
18	226.5	8.7	410	18	W12378
19	226	8.7	411	18	W12379
20	224	8.6	409	9	P81843
21	214	8.2	1010	19	W34989
22	213	8.2	467	17	W05731
23	213	8.2	467	17	W05382
24	211	8.1	411	18	W12381
25	209.5	8.0	412	18	W12380
26	193	7.4	499	14	R42122
27	189	7.2	551	18	W18790
28	186.5	7.2	352	20	Y08472
29	179.5	6.9	476	19	W37241
30	175	6.7	484	19	W43910
31	169	6.5	822	13	R26021
32	166	6.4	941	11	R07478
33	166	6.4	941	16	R77395
34	164	6.3	800	8	P70420
35	160	6.1	531	16	W01503
36	158.5	6.1	472	19	W37243
37	154.5	5.9	532	12	R13329
38	151.5	5.8	562	21	Y69508
39	150	5.8	357	16	R77394
40	143	5.5	521	17	R89927
41	142.5	5.4	358	19	W49874
42	142	5.4	358	19	W39262
43	129	4.9	1536	14	R41723
44	129	4.9	1536	14	R41725
45	129	4.9	1536	15	R63505

ALIGNMENTS

RESULT 1	
Y54122	
ID	Y54122 standard; Protein; 490 AA.
XX	
AC	Y54122;
DT	
XX	
DE	27-MAR-2000 (first entry)
XX	
XX	Amino acid sequence of a Bacillus mannase enzyme.
KW	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW	endo-1,4-mannanase; Bacillus sp. 163; galactomannan;
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW	galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW	printing paste; plant material degradation; recycled waste paper;
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;
KW	mannan-containing food; coffee extract; cleaning composition;
KW	machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW	contact lens; body-care composition; fabric softener; oil well drilling;
KW	subterranean formation fracture.
OS	
XX	
XX	Bacillus sp.
EH	
FT	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..30
FT	/note- "signal peptide"
FT	31..330
FT	Domain
FT	/note- "catalytic domain"
FT	Misc-difference 116
FT	/note- "encoded by CTT"
FT	Domain
FT	331..342
FT	/note- "linker"
FT	343..490
FT	Domain
FT	/note- "domain of unknown function"
XX	

Amino acid sequenc
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
P300-CelB fusion c
P300-CelB fusion c
Sequence of alkali
Terebinthacter end
Cellulase. Bacill
Bacillus cellulase
P300-CelB fusion c
P300-CelB fusion c
NK-1 cellulase. B
Corrected Bacillus
Actinomycete sp. 3
Heterodera glycine
Heterodera glycine
Alkaline cellulase
Cellulase. Bacill
Full length Bacill
Sequence encoded b
60 kD endoglucanase
Globochloa rostrat
Endoglucanase enco
Acidothermus cellu
Fragment of alkali
A. cellulolyticus
Bankia goudii glyco
A. cellulolyticus
High molecular wei
High molecular wei
Haemophilus high m

PN WO9964619-A2.
 XX 16-DEC-1999.
 XX
 XX 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 DR WPI: 2000-105891/09.
 DR N-PSDB; 245335.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions -
 XX
 PS Claim 1; Page 208-210; 242pp; English.
 XX
 XX The present sequence represents a mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
 CC of *Bacillus* sp. 1633. The mannanase hydrolyses galactomannans.
 CC Specifically, mannases hydrolyse 1,4-beta-D-mannosidic linkages in
 CC mannanase, galactomannans, glucomannans, and galactoglucomannans. The
 CC mannanase protein, or preparations containing it, are used to improve
 CC properties of cellulosic or synthetic fibres, yarn or (non)woven
 CC fabrics (removal of mannan-based sizes or printing pastes). They are
 CC also used to degrade or modify plant materials (particularly recycled
 CC waste paper, paper making pulps, or material containing guar or locust
 CC bean gums (thickeners), or to reduce viscosity of mannan-containing
 CC foods or feeds). The mannases are also used to process coffee
 CC extracts (to inhibit gel formation); in cleaning compositions (for
 CC machine washing of fabrics, as hard-surface cleaners, for hand or
 CC compositions) where they remove mannan-containing soils and prevent
 CC binding of some soils to cellulosics; and in fabric softeners. They
 CC can also be used in oil well drilling to fracture subterranean
 CC formations.
 CC
 CC Sequence 490 AA;
 XX
 XX

Query Match 100.0%; Score 2607; DB 21; Length 490;
 Best Local Similarity 100.0%; Pred. No. 1,9e-186;
 Matches 490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LNNGRKIFSTLSLILSLILSVSGTSTANNSGFYSGTLLVYANGNPFYMRINCH 60
 DB 1 lnngrkifstlsllslilsvsgtstannsgfyvgttlydangnpfymrglngh 60
 OY 61 AMYKQATTAETANTGANTVRIYVSDGGQWTKDDIHVRRLSLADNHLVAVPEYHD 120
 DB 61 amykqattalegtantgntvrrivsdggqwtkdihvrrllslaednhlvapevhd 120
 OY 121 ATGYDSISLNRADVWTENSGALIGKEDVTYIINANEMFGSWGEGDANADGKQAIPLRL 180
 DB 121 atgydsislnravdwtensgaligkedvtvlinanefwsgwedgawdgkqaiplr 180
 OY 181 NAGLHNTLMVDAAGGQFPGSIHDDGREVFADPPRNTMFSIHMTETAGNAGNAGVRRINID 240

DB 181 naglhtlmvdaagvgqfpgsihdygrevfadpprntmfsihmyetagnasgrrtnid 240
 OY 241 RVLNODLALVIGFCHRRITNGVDVDEATTMSYSEQRGVGLAMSKNGCPMEWYIDLNDW 300
 DB 241 rvlngdlalvlgfchrritngvdvdeattmsyseqvgvglawsvkngpeweyldlndw 300
 OY 301 AGNNLTAMGNTLVNPGVLRRTSRSTVFETGGSGSGSPPTIVDEBSMGWGTSSLSG 360
 DB 301 agnltamgntlvnpgvlrtrsrstvfetggsgsgspptivdegsmgwgtssls 360
 OY 361 GPWATWESSKSGSHSLKADIDLSNSQHYLVIONTSIQNSRIQATVKHANWGSVGNM 420
 DB 361 gpwatwessksgshskadidlsnsgqhylviontsiqnsriqatvkanwgsvgnm 420
 OY 421 TARKLVKTBGHTYTWSSGFVPINSGSGTTLSDLSNVONLSQVREIQFQASDSSGQT 480
 DB 421 tarklvktbghytwssgfvpinsgsgtllsdlsnvonlsqvreiqfqsads 480
 OY 481 STYIDNVAIE 490
 DB 481 styidnvai 490

RESULT 2
 Y44495
 ID Y44495 standard; Protein; 493 AA.
 AC Y44495;

27-MAR-2000 (first entry)

Bacillus agaradherens NCIMB 40482, Mannanase enzyme.

Mannanase: endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
 detergent composition; mid-branched anionic surfactant; washing;
 cosmetic stain; food stain.

Bacillus agaradherens.

Key Location/Qualifiers
 FT Peptide 1..32
 FT /label= Signal_peptide
 FT 33..343
 FT Protein /label= Mature_Mannanase

WO9964552-A1.

16-DEC-1999.

10-JUN-1998; 98WO-US12026.

10-JUN-1998; 98WO-US12026.

(PROC) PROCTER & GAMBLE CO.

Bettiol JP, Thoen CAJK;

WPI: 2000-116536/10.

N-PSDB; Z29845.

Detergent composition for removing greasy stains such as cosmetics,
 food stains and body soils -

Disclosure; Page 102; 113pp; English.

The present sequence is the *B. agaradherens* NCIMB 40482, alkaline
 mannanase enzyme. Mannanase shows maximum activity at pH ranging
 from 7.5-10.5. It can be used in a detergent composition along with a
 mid-branched anionic surfactant. The detergent composition may be used
 for washing purposes, soaking/pre-treatment of stained fabric, hard
 surface cleaning and for removal of cosmetic and/or food stains. This
 composition provides excellent cleaning effect at low temperature.

Query Match	65.68; Score 1711; DB 21; Length 468;
100%	100%

05-MAR-1999; 99DK-0000306.
05-MAR-1999; 99DK-0000306.
05-MAR-1999; 99DK-0000306.

23-DEC-1998; 98DK-0001725.

23-DEC-1998; 98DK-0001725.

```

PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Kauplien MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
DR WPI: 2000-105891/09.
DR N-PSDB: 245338.
XX
PT New mannanses for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions -
XX
PS Example 5; Page 215-216; 242pp; English.
XX
CC The present sequence represents a mannanase (also known as mannan
CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The present mannanase is a synthetic variant of the mannanase of
CC Y54123, in which the C-terminus of the protein was changed due to
CC design of a lower PCR primer used for amplification. The mannanase
CC hydrolyses galactomannans. Specifically, mannanses hydrolyse
CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanses
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC cellulosics; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.
CC
XX Sequence 468 AA;
SQ
Query Match 65.68; Score 1711; DB 21; Length 468;
Best Local Similarity 69.3%; Pred. No. 9,2e-120;
Matches 314; Conservative 60; Mismatches 75; Indels 4; Gaps 3;
XX
YY 11 ITLSILLASSILFVSGSTANANGSEYVSGTIVDANGNPFVNRGINGHVYKQARTTA 70
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
12 |tcllslsvgmgl-tspsaasgfyvdgntcllydaangpvtvmgindghayktaasta 70
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YY 71 IEGIANTGANTVRIYLSGSGOMTKDDITVNLISLADNHLVAPFEVHATGYDSIASL 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 71 |palaegantlrlvledeggyekddldtlrevlelaeqnmvavervhdatgrtsrdl 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YY 131 NRAVDYIEMRSALIGKEDVYIINANFEGSGWEGDANADGYKQALPRLRNAGLHNTLMV 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 131 nravdyltemkdallgkedvylinaaneygswdgsawadyldvlpkrlrdegldhltlmv 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YY 191 DAAGMGEPDSIHVGRFVNADPQRNMFSTIHMYEAGNAGNSQVTRINDRVLNDLALV 250
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 191 daagmgypgsihvgqvfnadpkrntmfshmyeagdantrvsnldrvldqdlav 250
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YY 251 IGEFGHRTNDVDENATIMSYEQRGVGLAMSWKNGKNGPEYIDLSDMAGNNNTANGN 310
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 251 lgefghtndvdenatimsyseqrgvglamswkngkngpewyldlsdewgghltdvgn 310
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YY 311 TIYNGPFGRLKRTSLRSLVFT--GGGSDGCTSPPTIYDFEGSGMQGTGSSLGCPMAVTEW 368
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 311 tiyngpfglkrtslrvft--ggsgdgcgtppti ydfegsgmqgtgsslgcpmavtew 368
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YY 369 SSKGSHSLKADIOIOLSSNGHILHYQNTSLQONSRIOATVHANAGSVCNGCTALYKXT 428
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 371 gasgnyslkadvnltstnseshelysegrnlhgyqlnatvtrhanwgnpngmarlykt 430
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YY 429 GHGTYWISGSFVPINGS-SGTTSLIDLSNVQNL 460
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 431 gsdylwshgspfftrlnssngtltlsfdlnlnteni 463
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 6
ID Y54123 standard; Protein; 476 AA.
XX Y54123;
AC Y54123;
XX
DT 27-MAR-2000 (first entry)
XX
DE A mannanase-linker-cellulose binding domain fusion protein.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;
KW 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW printing paste; plant material degradation; recycled waste paper;
KW paper making pulp; guar; locust bean gum; thickener; viscosity;
KW mannan-containing food; coffee extract; cleaning composition;
KW machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW contact lens; body-care composition; fabric softener; oil well drilling;
KW subterranean formation fracture; cellulose binding domain.
XX
OS Synthetic.
OS Bacillus sp.
OS Clostridium thermocellum.
XX
PN W09964619-A2.
PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Kauplien MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
DR WPI: 2000-105891/09.
DR N-PSDB: 245336.
XX
PT New mannanses for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions -
XX
PS Example 4; Page 211-212; 242pp; English.
XX
CC The present sequence represents a mannanase-linker-cellulose binding
CC domain fusion protein. Mannanase (also known as mannan
CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
CC hydrolyses galactomannans. Specifically, mannanses hydrolyse
CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
CC and galactoglucomannans. The mannanase protein, or preparations
CC containing it, are used to improve properties of cellulosic or

```

CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannans
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC celluloses; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.

CC Sequence 476 AA:

Query Match 62.9%; Score 1640; DB 21; Length 476;
 Best Local Similarity 68.6%; Pred. No. 1,8e-114;
 Matches 334; Conservative 30; Mismatches 73; Indels 50; Gaps 9;

QY 32 ANSGFVSGTTLTDANGNPFVWKGTHGAWTKDQATTAIEGANTGANTVIVLSDGQ 91
 DB 1 ansqfvsqtltydangnpfvmgthgnawkqdcatalegiantgtvrvlsdgsq 60
 QY 92 WTKDDIHTVKNLISLAEDNHLVAVPEVHDATGYDSTASINRAVDYIEMKRSALICKEDTV 151
 DB 61 wtkddihtrnlislaednhlvavpevhdatgydaslnravdyiemrsalickedtv 120
 QY 152 ITIANEMTGSWGDWADAGYQAIPTLRNAGNHTLMDAGMGQFQSHIDYGREYFN 211
 DB 121 itlanewtgswegdawadgykqaiptlrnaglnhltmdaagwqfqpshldygreyn 180
 QY 212 ADQRNTMPSIHMYEYAGNAGSVRTNIDRVLNQDLALVIGFGHHTNGDVEATIMSY 271
 DB 181 adqrntmfshmyeyagnasqvrtnidrvlnqdlalvigfghhtngdveatimay 240
 QY 272 SEQRGVMTLMSWKGKPEWEYIDLSDNAGNLTANGTIVGPRGLAETSTSLVFTG 331
 DB 241 seqrgvmtlmswkgkpeweyidlndnagnltawgnltvpgylretesrlvfta 300
 QY 332 G-----GSDGGRPTTLVDFGSMQGTSSLSGFPNAVTEMSSKSHSLKADIQLS 384
 DB 301 speptpeptantcvsqnlkvefynspstlnln-pqfvtntlgasa-----ldlak 352
 QY 385 NSOHTLVIONTSIQNSRIQATVKHANGSVNGMTARLYVTGAGTYTSSFTPIG 444
 DB 353 ltrlyvtvqgkqctfwdhaalngn-gsy-ngltanv-----kglfkms 399
 QY 445 SSG-----TTLISDL-----SNVQ-----NLQVREIGVFOGASDSSGQTSI 482
 DB 400 stnadtylelstfgtglepahnvlggrfakndwnyrtgndysfkrsrqfwewdqvta 459
 QY 483 YIDNVIV 489
 DB 460 ylngvlyv 466
 RESULT 7
 Y54127 Y54127 standard; Protein; 331 AA.
 AC Y54127;
 DT 27-MAR-2000 (first entry)
 DE Amino acid sequence of a Bacillus mannanase enzyme.
 XX Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;
 XX endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 XX mannan; galactomannan; galactomannan; galactoglucomannan; cellulosic fibre;
 XX synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;

KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.

XX Bacillus sp.

FX Key location/Qualifiers
 FT Peptide 1..32 "signal peptide"
 FT Domain 33..331
 FT Domain /note= "catalytic domain"

XX MO0964619-A2.

XX 16-DEC-1999.

XX 10-JUN-1999;

XX 99WO-DK00314.

XX 10-JUN-1998;

XX 20-OCT-1998;

XX 20-OCT-1998;

XX 28-OCT-1998;

XX 28-OCT-1998;

XX 23-DEC-1998;

XX 05-MAR-1999;

XX 05-MAR-1999;

XX 05-MAR-1999;

XX 09-MAR-1999;

XX 10-MAR-1999;

XX 11-MAR-1999;

XX (NOVO) NOVO-NORDISK AS.

XX Kaupinen MS, Schnelein M, Schmitt K, Andersen LN, Bjornvad ME;

XX WPI: 2000-105891/09.

XX N-PSDB: Z45340.

XX New mannases for treatment of textiles, plant material and coffee

XX extract, and in cleaning compositions

XX Claim 35; Page 220-221; 242pp; English.

XX The present sequence represents a Bacillus mannanase (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to celluloses; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.

XX Sequence 331 AA:

Query Match 44.2%; Score 1152; DB 21; Length 331;

Best Local Similarity 65.0%; Pred. No. 2.4e-78;

Matches 208; Conservative 54; Mismatches 58; Indels 0; Gaps 0;

QY 10 SITLSLASSILFVSGTSTANNSGFYVSGTTLTDANGNPFVWKGTHGAWTKDQAT 69

```

Db 11 alicsllfagvlnatssgaeehbsgfnvngtllfydangnpfvmrglnghwfxgelet 70
QY 70 AIEGANTGANTVRIYVSDGGQMTKODIHTVYVNLISLAEQNLVAVPEYHATGYSIAS 129
Db 71 smrgtsqgntllrvlndangrkwqkddrmvavslaeqbmqlavlewhdatsnntsd 130
QY 130 LNRADVYIEMRSLIGEDPVIINIANMFGSMGEGDADQYKQAIPLRARG:GNHTLM 189
Db 131 lgaadvylmndvlgqkredvllnlanevgaqdgawaqgallqrlmagslhtlm 190
QY 190 VDAAGGQFPOSIDHYGREGVENADPQRTMFSIHMYETAGNAGSQVRTNIDRYLNDLAL 249
Db 191 vdaegygygpgsvvdygqevlnedpqrntmfsvhmyeyagggancvrrnldslsqnla 250
QY 250 VIGEGHRTNGDVEDATIMSYSEQRYGWTAMSKNGSPWEYIDLSDMAGNNLTAWG 309
Db 251 vlgefghvhyddvdedctllsygqgrnvglawshvgnasegyldlndfagrnltwvg 310
QY 310 NTIVNGPYGLRSTRSLTYF 329
Db 311 drlvgnglrlqtskrsvf 330

RESULT 8
Y54129
ID Y54129 standard; Protein; 369 AA.
XX
AC Y54129;
XX
DT 27-MAR-2000 (first entry)
XX
DE Amino acid sequence of a Bacillus sp. AA349 mannanase enzyme.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
    endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
    mannan; galactomannan; galactoglucomannan; cellulostic fibre;
    synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
    plant material degradation; recycled waste paper; paper making pulp;
    guar; locust bean gum; mannan-containing food; coffee extract;
    cleaning composition; machine washing; hard-surface cleaner;
    dishwashing; oral; dental; contact lens; body-care composition;
    fabric softener; oil well drilling; subterranean formation fracture.
XX
OS Bacillus sp.
XX
FH Key
FT Peptide
FT /note= "signal peptide"
FT Domain
FT /note= "N-terminal domain of unknown function"
FT /note= "catalytic domain"
XX
PN W09964619-A2.
PD 16-DEC-1999.
PF 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.

```

```

XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
DR WPI: 2000-105891/09.
XX
DR N-PSDB: Z45342.
XX
PT New mannanases for treatment of textiles, plant material and coffee
    extract, and in cleaning compositions
XX
PS Claim 35; Page 224-225; 242pp; English.
XX
CC The present sequence represents a mannanase enzyme (also known as
    mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
    The mannanase hydrolyses galactomannans. Specifically, mannanases
    hydrolyse 1,4-beta-D-mannosidic linkages in mannan, galactomannan,
    glucomannans, and galactoglucomannans. The mannanase protein, or
    preparations containing it, are used to improve properties of cellulosic
    or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
    sizes or printing pastes). They are also used to degrade or modify
    plant materials (particularly recycled waste paper, paper making pulps,
    or material containing guar or locust bean gums (thickeners), or to
    reduce viscosity of mannan-containing foods or feeds). The mannanases
    are also used to process coffee extracts (to inhibit gel formation); in
    cleaning compositions (for machine washing of fabrics, as hard-surface
    cleaners, for hand or machine dishwashing, also in oral, dental, contact
    lens or body-care compositions) where they remove mannan-containing
    soils and prevent binding of some soils to celluloses; and in fabric
    softeners. They can also be used in oil well drilling to fracture
    subterranean formations.
XX
SO Sequence 369 AA;

Query Match 41.2%; Score 1075; DB 21; Length 369;
Best Local Similarity 66.2%; Pred. No. 1.5e-72;
Matches 196; Conservative 37; Mismatches 63; Indels 0; Gaps 0;

QY 34 SGFYSGTITLYANGNPVNRGINSHGAWYKQDATTAEIGANTGANTVRIYVSDGGQMT 93
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 ngfyldgtllydangkrpvmrglnghvwykphstamealadfganslrvlsdsgqvt 131

QY 94 KODIHTVNLISLAEQNLVAVPEYHATGYSIASLNRADVYIEMRSLIGEDPVIIT 153
    |||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 132 kddvdevakllslaeqhsalvaalevhdalgytddlepllktvdyvlelxdaligkdxvll 191

QY 134 NIANMFGSMGEGDADQYKQAIPLRARG:GNHTLMVDAAGGQFPOSIDHYGREGVENAD 213
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 192 nlsnewfsgvseggwadykkaipllrleaqlkhtlmvdaagvgdfrslhcxglevnsd 251

QY 214 FORNTMFSIHMYETAGNAGSQVRTNIDRYLNDLALVIGEFGRNTNDVEDATIMSYSE 273
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 252 plknmtfshyewaaagpqqvkdldgvlknlavvlgelfghhygvdavdltlshse 311

QY 274 ORGVGWTAMSKNGSPWEYIDLSDMAGNNLTAWGNTIVNGPYGLRSTRSLTYF 329
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 312 kydvglawshvgnasegyldlrdtsqgtlvtewgerlvpnglketseivsvy 367

RESULT 9
Y54132
ID Y54132 standard; Protein; 320 AA.
XX
AC Y54132;
XX
DT 27-MAR-2000 (first entry)
XX
DE Amino acid sequence of a Bacillus sp. mannanase enzyme.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
    endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
    mannan; galactomannan; glucomannan; galactoglucomannan; cellulostic fibre;

```

synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 plant material degradation; recycled waste paper; paper making pulp;
 guar; locust bean gum; mannan-containing food; coffee extract;
 cleaning composition; machine washing; hard-surface cleaner;
 dishwashing; oral; dental; contact lens; body-care composition;
 fabric softener; oil well drilling; subterranean formation fracture.
 Bacillus sp.
 Key Location/Qualifiers
 Peptide 1..28
 /note= "signal peptide"
 FT 29..320
 FT Domain /note= "catalytic domain"
 W0964619-A2.
 16-DEC-1999.
 10-JUN-1999; 99WO-DK00314.
 10-JUN-1998; 98US-0111256.
 20-OCT-1998; 98DK-0001340.
 20-OCT-1998; 98DK-0001341.
 28-OCT-1998; 98US-0105970.
 28-OCT-1998; 98US-0106054.
 23-DEC-1998; 98DK-0001725.
 05-MAR-1999; 99DK-0000306.
 05-MAR-1999; 99DK-0000307.
 05-MAR-1999; 99DK-0000308.
 05-MAR-1999; 99DK-0000309.
 09-MAR-1999; 99US-0123543.
 10-MAR-1999; 99US-0123623.
 11-MAR-1999; 99US-0123641.
 11-MAR-1999; 99US-0123642.
 (NOVO) NOVO-NORDISK AS.
 Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 WPI: 2000-105891/09.
 N-PSDB: 245345.
 New mannases for treatment of textiles, plant material and coffee
 extract, and in cleaning compositions
 Claim 35; Page 228-229; 242pp; English.
 The present sequence represents a mannanase enzyme (also known as
 mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to celluloses; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 SO Sequence 320 AA;

Query Match 36.7%; Score 957.5; DB 21; Length 320;
 Best Local Similarity 58.4%; Pred. No. 7.1e-64;
 Matches 181; Conservative 42; Mismatches 86; Indels 1; Gaps 1;

QY 16 LLAASILFVGSTANANGSFYVSGTTLVYDANGNPVMEGINCHAMVYKQATTAIEGIA 75
 Db 12 LVVVIVIVYSSGLAEEGSGFHVYKTEILDKNGDPYVMYGVNHGVSFKGDLSEALPAIA 71
 QY 76 NTGANTRYIYLSGGGOWTKDIDHTRNLISLAEDNHLVAVPEVHDANGYSIASLNAVD 135
 Db 72 ETGANLTVIVLSNGGQWKKDAEELARVIAETGYLTLEVHDATGSDNPDDLDAVD 131
 QY 136 YWIMERSALIGKEDTVIINIANEFWSGMEGDAMADGYKQAIPIRLNGLNHTLVADAGW 195
 Db 132 YWIMAEAVLKGTEDEVILINLANEYGAWSADVAEYAGAIPIRLSEGLHTLLVDAEW 191
 QY 196 GQFQSHIDHREYFNADPQNRNMFSLHMYEXAGNSQVNRINDRYLNDLALVIEFG 255
 Db 192 GQYASLHERGADVFAEDPIKNTMFSIMHYEGADRATVSENIDYLAENLAVVIEFG 251
 QY 256 HRHTNGVDVDEATIMSYSEORGVGLTAMSKGNGEPWEYLDISNDWAGNLTAMGNTVING 315
 Db 252 HRHDGVDVDEALAYTAERGVLAVSYGNSGVVEYLDITGPGS-PLTSGSERIVYG 310
 QY 316 PYGLRSTSL 325
 Db 311 EMGLKVIDHL 320
 RESULT 10
 Y54130
 ID Y54130 standard; Protein; 305 AA.
 XX Y54130;
 AC Y54130;
 XX 27-MAR-2000 (first entry)
 DE Amino acid sequence of a partial Bacillus sp. mannanase enzyme.
 XX
 KW Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 XX
 OS Bacillus sp.
 XX
 PN W0964619-A2.
 XX 16-DEC-1999.
 PD 10-JUN-1999; 99WO-DK00314.
 XX 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 11-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 PA (NOVO) NOVO-NORDISK AS.
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX


```

Db      72  etgantvrlvlsngqwekdsselarvlatetyglvtlevhdatsgndpddkavd 131
Qy      136  WIENRSALIGREDYVIINIANEFWSGEMADGYQALPRLNAGLNTLMDA 192
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      132  ylemadvlkgfedrvlinlanewyawsdlwakayagaprlrsaglanllida 188

RESULT  12
P91007 standard; protein; 487 AA.
ID      P91007
AC      P91007;
XX      13-MAR-1990 (first entry)
DE      Beta-mannase of alkalophilic Bacillus.
KW      Escherichia coli.
OS      Bacillus sp.
XX      JP01228477-A.
XX      12-SEP-1989.
XX      09-MAR-1988; 88JP-0053774.
XX      PF      03-MAR-1988; 88JP-0053774.
XX      PR      (SHK7) SHINGIUTUSU KAIHATSU.
XX      PA      WPI; 1989-312223/43.
XX      DR      N-PADB; N91621.
XX      PT      beta-mannase gene deoxyribonucleic acid
XX      PT      - obtd. by culturing transformed Escherichia coli
XX      PT      contg. recombinant plasmid
XX      PS      Claim 1; Fig 3; 10pp; Japanese.
XX      CC      E. coli transformants produce beta-mannase efficiently. It is obtd.
XX      CC      from Bacillus sp. PERM P-8856, -8857 and -8858.
XX      SQ      Sequence 487 AA;

Query Match      16.1%; Score 419; DB 10; Length 487;
Best Local Similarity 27.0%; Pred. No. 1.6e-23;
Matches 142; Conservative 78; Mismatches 203; Indels 102; Gaps 22;

Qy      25  SGSTSTANSGFYSGTTLTDAN-----GNPFVARGINHG-----AMTKD---QATTAIE 72
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      4  ngaalspananqtcknyvslanlpnksnkrrvsghtgydsdclawlkqcarelqkmp 63
Qy      73  GIANTGANT--VRIVLSDGGQWTRDDJHTVNLISLAEDNHLVAV-----PEVHDATGY 124
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      64  gllscdyknwclrlvvdqslsgca-----qellnfmgqglvrlsvhmpnpgfhsgeuy 118
Qy      125  DSIALNRNAVY-----WIENRSALIGKED-----TVIINANEFWSGEMDA 167
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      119  ktlipstgfgqlntgrttegrrrkmdlkmadgldelqngvtvlfrplhemngew---f 175
Qy      168  W--ADGKYQALPRLNAGLN-----HTLMVDAAGNQGFPQSIHDYGREVFNDPQ 215
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      176  wmgsegynqfdqtranyaisawrdmyqyftnerkinnllwyspdyrdh---vtsyypg 232
Qy      216  RNTPESSIMYEGAGNAGQVNTIDRVLNODLALVYEGFGRHT--NGDVDEATIMSYSEQ 274
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      223  anyvdvaldery-hpdhslidqynrmaldkpfafaelgspemagfdysnlgaklq 291
Qy      275  ---RGVGMILANS--WK--GNGPEWEYLDLSNDWAGNNITLNGNTIVNGPYGLRETSRLST 327
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      292  kyptrvyflavndkvspnngaw---dlfnd-----sw 322

```

```

Qy      328  VFTGGSGDGGRS--PTTLYDFEGSGMOGTGSSUS--GCPNATWESSKGSLSKADQLQS 384
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      323  vvnrgeldygsnpacvrllydfennllswagcelfdgppwslnewsanqfslkadvlg 382
Qy      385  NSQHYLVHVIQNTSLQONSRLQATYKXANWGSVGNNGWTARLYVTGTGYMSSGFPI 444
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      383  nsyh-lqktyvrnlssfknlslkxshswgnvgsgmlarlvfxkgsawrnwagfcqfag 441
Qy      445  SSGTTLSDLSNVONLSQVREIGVQFQASDSSGQTSIYIDNVIY 489
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      442  krtalstdltkvsnlhvzreigreykapansngkalyldhvtv 486

RESULT  13
Y54131
ID      Y54131 standard; Protein; 132 AA.
XX      Y54131;
XX      AC      Y54131;
XX      27-MAR-2000 (first entry)
DE      Amino acid sequence of a partial Bacillus mannanase enzyme.
XX      XX
XX      Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX      endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
XX      mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
XX      synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
XX      plant material degradation; recycled waste paper; paper making pulp;
XX      guar; locust bean gum; mannan-containing food; coffee extract;
XX      cleaning composition; machine washing; hard-surface cleaner;
XX      dishwashing; oral; dental; contact lens; body-care composition;
XX      fabric softener; oil well drilling; subterranean formation fracture.
XX      OS      Bacillus clausii.
XX      W09964619-A2.
XX      16-DEC-1999.
XX      PD      10-JUN-1999; 99WO-DK00314.
XX      PF      10-JUN-1998; 98US-0111256.
XX      PR      20-OCT-1998; 98DK-0001340.
XX      PR      20-OCT-1998; 98DK-0001341.
XX      PR      28-OCT-1998; 98US-0105970.
XX      PR      28-OCT-1998; 98US-0106054.
XX      PR      23-DEC-1998; 98DK-0001725.
XX      PR      05-MAR-1999; 99DK-0000306.
XX      PR      05-MAR-1999; 99DK-0000307.
XX      PR      05-MAR-1999; 99DK-0000308.
XX      PR      05-MAR-1999; 99DK-0000309.
XX      PR      09-MAR-1999; 99US-0123543.
XX      PR      10-MAR-1999; 99US-0123623.
XX      PR      10-MAR-1999; 99US-0123641.
XX      PR      11-MAR-1999; 99US-0123642.
XX      PA      (NOVO) NOVO-NORDISK AS.
XX      KAuppien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX      WPI; 2000-105891/09.
XX      DR      N-PADB; Z45344.
XX      PT      New mannanases for treatment of textiles, plant material and coffee
XX      PT      extract, and in cleaning compositions
XX      PS      Disclosure; Page 227; 242pp; English.
XX      CC      The present sequence represents a mannanase enzyme (also known as
XX      CC      mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
XX      CC      The mannanase hydrolyses galactomannans. Specifically, mannanases
XX      CC      hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,

```

CC This protein comprises the alkaline cellulase (endoglucanase) of
CC *Bacillus agaradhaerens* NCTMB 40482. It was expressed in *Bacillus*
CC subtilis PL306 transformants following PCR amplification (see
CC T74286-89) of *B. agaradhaerens* genomic DNA and ligation of the PCR
CC product into vector pN1581. DNA encoding the cellulose binding

PI Bjornvad ME, Rasmussen MD, Vind J, Von der Osten C;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:00:18 ; Search time 154.68 Seconds
(without alignments)
53.097 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607

Sequence: 1 LAMGFKKIFSTLSLILASS.....QSASDSCGCTSIIDNVIVE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/PCITUS_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229.5	8.8	400	2	US-08-713-298B-2
2	229.5	8.8	400	2	US-08-870-180B-2
3	229.5	8.8	400	3	US-08-814-052-4
4	229.5	8.8	400	3	US-08-812-829-4
5	229.5	8.8	462	2	US-08-870-180B-13
6	213	8.2	467	2	US-08-727-548-2
7	189	7.2	551	2	US-09-033-537A-1
8	174	6.7	430	2	US-08-924-440-2
9	160	6.1	531	2	US-07-862-588B-7
10	143	5.5	521	1	US-08-276-213-3
11	142	5.4	358	1	US-08-604-913B-11
12	129	4.9	1356	1	US-08-038-682-2
13	129	4.9	1536	1	US-08-302-832-2
14	129	4.9	1536	2	US-08-530-198-2
15	129	4.9	1536	2	US-08-469-880-2
16	129	4.9	1536	2	US-08-728-470-2
17	129	4.9	1536	2	US-08-617-697-2
18	123.5	4.7	680	3	US-08-947-965-77
19	116.5	4.5	983	3	US-08-164-292B-26
20	116.5	4.5	983	3	US-08-845-623-26
21	116.5	4.5	983	3	US-08-815-927-26
22	112.5	4.3	480	1	US-08-672-571A-1
23	111	4.3	574	3	US-08-732-433-1
24	109	4.2	735	2	US-08-313-185-48
25	109	4.2	735	2	US-08-459-499-9
26	108	4.1	1545	4	PCT-US95-10661A-4
27	107.5	4.1	666	2	US-08-737-716-14
28	107.5	4.1	1222	2	US-08-682-517-15

29	107.5	4.1	1252	2	US-08-682-517-9	Sequence 9, Appl1
30	106.5	4.1	750	5	5457037-3	Patent No. 5457037
31	106	4.1	1338	2	US-08-728-470-9	Sequence 9, Appl1
32	106	4.1	1599	2	US-08-617-697-9	Sequence 9, Appl1
33	105.5	4.0	751	5	5457037-5	Patent No. 5457037
34	105	4.0	385	2	US-08-387-942C-26	Sequence 26, Appl1
35	105	4.0	654	1	US-08-392-828C-2	Sequence 2, Appl1
36	105	4.0	654	3	US-09-330-945-2	Sequence 2, Appl1
37	105	4.0	997	2	US-08-387-942C-4	Sequence 4, Appl1
38	104	4.0	358	2	US-08-853-659A-45	Sequence 12, Appl1
39	104	4.0	735	2	US-08-459-499-12	Sequence 45, Appl1
40	103.5	4.0	553	2	US-08-661-052-16	Sequence 16, Appl1
41	103	4.0	434	2	US-08-989-925-1	Sequence 1, Appl1
42	103	4.0	1702	4	PCT-US95-10661A-5	Sequence 5, Appl1
43	102.5	3.9	623	3	US-09-041-991A-6	Sequence 6, Appl1
44	101.5	3.9	566	2	US-07-862-588B-4	Sequence 4, Appl1
45	101.5	3.9	928	1	US-08-474-140-11	Sequence 11, Appl1

ALIGNMENTS

RESULT 1
US-08-713-298B-2
; Sequence 2, Application US/08713298B
; Patent No. 5922586
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Dammann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg rd-Frantzen, Henrik
; APPLICANT: Sch lein, Martin
; APPLICANT: J rgensen, Per
; APPLICANT: Bjergvad, Mads
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS: 11
; ADDRESS: No. 5922586 of No. 5922586disk of No. 5922586th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713, 298B
; FILING DATE: 13-SEPT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 3794.424-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 867-0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-713-298B-2

Query Match 8.8%; Score 229.5; DB 2; Length 400;
Best Local Similarity 23.4%; Pred. No. 5.8e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25 (Epo)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08,870,180B
 FILING DATE: 6-June-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rozeck, Carol E.
 REGISTRATION NUMBER: 36,993
 REFERENCE/DOCKET NUMBER: 3794.434-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 867 0298
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 400 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

[illegible]

GENERAL INFORMATION:
APPLICANT: von der Osten, Claus
APPLICANT: Chet'ry, Joel R.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Vind, Jesper
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOLLING
OR STAINS FROM CELLULOIC FABRIC
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:

ADDRESSSEE: NO. 60157830 NO. 601578341sk
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/814,052
FILING DATE: 06-MAR-1997
CLASSIFICATION: 510
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4684.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-814-052-4

Query Match 8.8%; Score 229.5; DB 3; Length 400;
Best Local Similarity 23.4%; Pred. No. 5.8e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

QY 6 KRFISITLILASSILFVSGSTANANS-----GFVYSGTLLDANGNPVWNGI-NH 58
DB 2 KRTITLIVLMTVALFSGITTAADNDVSVEHGOQLSNGELVNERGEQYOLKGMSSH 61
QY 59 GHAVY-----KQATTAEGIANGTAVRVL--SDGGQWTKDDI-HTVRNL 103
DB 62 GLOWGQFVNESMKMLRD-----DWGINFRAMNTSSGGYIDDPVKKKYEKA 111
QY 104 ISLADNHLVAVPEVHDATGDSIASLNRAVDYWTMSALIGKEDTYIINIANEFQSW 163
DB 112 VEAALDILYIYIIDHILSDNDPNITKEAKDFDEM--SELGDPNVIYEIANEPNCS- 169
QY 164 EGDADADGK-----QAPRLRAGLNHTLMDVADAGQFPOSIDHYGREVENADPQRNTM 219
DB 170 -DVTGNQIKPYAEVPIPIIRNNDPNITIV--GTGWSQDVH--HADNQLADP--NVM 222
QY 220 FSIHMYEYAGNASQVRNIDVNLQDIALVIGFGRHTNGD---VDEATI-MSYSQ 274
DB 223 YAFHF--YAGTGGVLRQVYALDQGAIFVSENGTSATGDGVFLDEAOWIDFDE 280
QY 275 RGVGVLANS-----WKNGPWEYIDLSDNAGNNTLAVNGTIVNGPYGLRETS 323
DB 281 RNLSTWANSILFKDESSAALMPGAMP-----TCGWTAEALSPSGTIVRE--KIRESA 330
QY 324 RUSTYFTGGSGSGTSPITLYDFESSMOGWGSSL-----SGGPAVATWMS 369
DB 331 SIIPSDPTPPSDPGEPDPTPPSDPGEYPAMPNQIYINELIYVHNQOLWQAKWT 384

RESULT 4
US-08-812-829-4
Sequence 4, Application US/08812829
Patent No. 6017751
GENERAL INFORMATION:
APPLICANT: von Oer Osten, Claus
APPLICANT: Bjornvad, Mads E.
APPLICANT: Wind, Jesper
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
NUMBER OF INVENTION: OR STRAINS FROM CELLULOSE FABRIC
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 60177510 No. 6017751disk of No. 6017751th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812.829
FILING DATE: 06-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4690,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-812-829-4

Query Match 8.8%; Score 229.5; DB 3; Length 400;
Best Local Similarity 23.4%; Pred. No. 5.8e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

QY 6 KRFISITLILASSILFVSGSTANANS-----GFVYSGTLLDANGNPVWNGI-NH 58
DB 2 KRTITLIVLMTVALFSGITTAADNDVSVEHGOQLSNGELVNERGEQYOLKGMSSH 61
QY 59 GHAVY-----KQATTAEGIANGTAVRVL--SDGGQWTKDDI-HTVRNL 103
DB 62 GLOWGQFVNESMKMLRD-----DWGINFRAMNTSSGGYIDDPVKKKYEKA 111
QY 104 ISLADNHLVAVPEVHDATGDSIASLNRAVDYWTMSALIGKEDTYIINIANEFQSW 163
DB 112 VEAALDILYIYIIDHILSDNDPNITKEAKDFDEM--SELGDPNVIYEIANEPNCS- 169
QY 164 EGDADADGK-----QAPRLRAGLNHTLMDVADAGQFPOSIDHYGREVENADPQRNTM 219
DB 170 -DVTGNQIKPYAEVPIPIIRNNDPNITIV--GTGWSQDVH--HADNQLADP--NVM 222
QY 220 FSIHMYEYAGNASQVRNIDVNLQDIALVIGFGRHTNGD---VDEATI-MSYSQ 274
DB 223 YAFHF--YAGTGGVLRQVYALDQGAIFVSENGTSATGDGVFLDEAOWIDFDE 280
QY 275 RGVGVLANS-----WKNGPWEYIDLSDNAGNNTLAVNGTIVNGPYGLRETS 323
DB 281 RNLSTWANSILFKDESSAALMPGAMP-----TCGWTAEALSPSGTIVRE--KIRESA 330
QY 324 RUSTYFTGGSGSGTSPITLYDFESSMOGWGSSL-----SGGPAVATWMS 369
DB 331 SIIPSDPTPPSDPGEPDPTPPSDPGEYPAMPNQIYINELIYVHNQOLWQAKWT 384

RESULT 5
US-08-870-180B-13
Sequence 13, Application US/08870180B
Patent No. 5945327
GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dammann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Bisg id-Frantzen, Henrik
APPLICANT: Sch lehn, Martin
APPLICANT: J rgensen, Per
APPLICANT: Bjornvad, Mads
TITLE OF INVENTION: DNA Constructs and Methods of Producing
NUMBER OF INVENTION: Cellulytic Enzymes
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59453270 No. 5945327disk of No. 5945327th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

FILING DATE: 02-MAR-1998
 CLASSIFICATION: 008
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 0993/95
 FILING DATE: 08-SEP-1995
 APPLICATION NUMBER: PCT/DK96/00364
 FILING DATE: 03-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Green, Reza
 REGISTRATION NUMBER: 38,475
 REFERENCE/DOCKET NUMBER: 4492, 204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 551 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-033-537A-1

Query Match 7.2%; Score 189; DB 2; Length 551;
 Best Local Similarity 21.3%; Pred. No. 2.9e-08;
 Matches 108; Conservative 72; Mismatches 181; Indels 146; Gaps 24;

38 VSGTLLYDANGNPFYMGRI-NHGAWYDQATTAEGIANFGANTVRIYVLSGGGWTDD 96
 11 VQGNQVVGSGQAVQVVGMSHGLQWYGN-----FYNKSLSQWMDN 52
 97 -IHVRLISLAEDNHLVAVPEVDATGYSIASLNA-----VDYNT----- 138
 53 WGINFRAAMTAEDGY-ITDPYANKKVEAVQASIDGLYIIDWHILSDGNPTTYRAQ 111
 139 -----EMSAIIGKEDVTIINANEMFGSWEGDANAD--GY-KQAIPLRLNAGLNTLAV 190
 112 SKAFQEQATLYGNTPNYIETIANEPNGN---VSNADKSTAEVITLIRAIIDPGVIV 168
 191 DAAGGQFPOSIDHYGREVFNADPORNTFSIHMYEYAGNASOVRTNIDRYLNODLALV 250
 169 GSPFWSQ-----DIHLAADNPVSHSNVYALHF--YSGTHQGFRLDRITTYAMNKGALIF 220
 251 IGEGRHTND-----VDEATIMSISQRGVGLAWS-----WNGGPEWEYL 294
 221 VTEGTSIDASGNGGPFPOSKEWIDFLNARKISWVNMISLADKVENSAALMPGASP----- 275
 295 DLSMDMAGNNLTJANGNTIVNGPYGLRETSRLSTVFTGGSDGCT--SPTLLYDEGSMQ 351
 276 --TGGWDAQLSSEGRVY-----RQDIROA--TGGSGNPTIAPAPNTLSATAGNAQ 323
 352 -GWTGSSLSGGPMAVTESSKSHSLKADIOSNSOYHLVHTONTLSQNSRIQATYKH 410
 324 VSLTMNAVSGAT-----SYTVK-----RATISG 346
 411 ANMSVSGMGRARLKYTG--HGTYWY-----SGSFVINGSSGTTLSLDSNVONLSQV 463
 347 GPYNVATGVTATSTYNGTITNGTITTYVVSANS---AGSSANSQAASATPAGGAST 402
 464 REIGVQFO-----SASDSSGQTSIYIDN 486
 403 GNLVQVYKVGDTSATDNOMKPSFNIN 429

RESULT 8
 US-08-924-440-2
 Sequence 2, Application US/08924440
 Patent No. 5871550
 GENERAL INFORMATION:
 APPLICANT: Frits et al.
 TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
 NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genencor International, Inc.
 STREET: 925 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/924,440
 FILING DATE: August 27, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Christopher L. Stone
 REGISTRATION NUMBER: 35,696
 REFERENCE/DOCKET NUMBER: GC388
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 846-6504
 TELEFAX: (650) 846-6504
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 430 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-924-440-2

Query Match 6.7%; Score 174; DB 2; Length 430;
 Best Local Similarity 24.7%; Pred. No. 4e-07;
 Matches 81; Conservative 49; Mismatches 130; Indels 68; Gaps 19;

26 GTSFNANSG-FYVSGTLLYDANGNPFYMGRI-NHGAWYDQAT-TAIEGTA-NTGANT 81
 124 GTGPFVERYGQVQCGIDLCDEHGNPVQLRGMSTHGIDFCHCLDSSLDLALYPMKADI 183
 82 VRI---VLSDGQWT---KDDIHVRLISLAEDNHLVAVPEVDATGYSIASLNAV 134
 184 IRLSMTIODEGTEINPGFTDRH---QLIDMATAKGLYIVDWIILLPGDPHYVLDRAK 240
 135 DWIEMRSLIGKEDVTIINANEMFGSWEGDANAD--GY-KQAIPLRLNAGLNTLAVD 191
 241 TFFAEIARHASKIN-VIETLANE-----PNGVSASIKSTAEVITPIRQDDPSVITVG 295
 192 AAGWGQFPOSIDHYGREVFNADP--QNTMFSIHMYEYAGNASOVRTNIDRYLNODLAL 249
 296 TRGWSLSGV--EGSGPAEIANPVAASINMYAFHY-----AASHRDYINLALREASEL 348
 250 ---VIGERGHRTND-----VDEATIMSISQRGVGLAWSK----- 285
 349 PPFVETFERGTETTYGDGANDFQMDR--YIDLMARKIGWTKMNSDDPFGAGVFPQPTC 406
 286 -GNGPEWEYLDLSNDMAGNNLTJANGNTI 312
 407 ASGGP-----WSSGSLKASGQWV 424

RESULT 9
 US-07-862-588B-7
 Sequence 7, Application US/07862588B
 Patent No. 5916796
 GENERAL INFORMATION:
 APPLICANT: Joergensen, Per Linnae
 APPLICANT: Sch lein, Martin
 APPLICANT: Hansen, Christian
 TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
 NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
 STREET: 405 Lexington Avenue, 62nd floor
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10017

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/862,588B

FILING DATE: 19920727
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: DK 164/90

FILING DATE: 19-JAN-1990

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/DK91/00013

FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:
 NAME: Nelson, Steve T. / Lamberts, Elias J.

REGISTRATION NUMBER: 30,335 / 33,728
 REFERENCE/DOCKET NUMBER: 3425.204-US

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
 LENGTH: 531 amino acids

TYPE: AMINO ACID
 TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-862-588B-7

Query Match 6.1%; Score 160; DB 2; Length 531;
 Best Local Similarity 20.6%; Pred. No. 8.9e-06;
 Matches 109; Conservative 73; Mismatches 214; Indels 137; Gaps 24;

11 ITTSLIASILFVSGSTANANG-----FYVSGTLYDANGNPFVNGI-NHGAWTK 64
 12 LVLTWVLMGLLPVGAPEKTAAPVFGQAKQGNQVQSGQAVQVQMSHGLQWVG 71
 65 DQATTAIEGIANGTANTVRLVLDGGQWTD--HTVRNLISLADNHLVAVPEVDAT 122
 72 N-----FVNRKSLQWMDNNGINVFRAAMTSEGGY-ITDPSYKNTV 112
 123 GYDSIASLNR-----VDYWT-----ENRSALIGKEDVTIINIANEWG 161
 113 KEAVQASIDIALYIIDWHLSDGNPTTYAQSFAFFQEMATLYKNTPNYIELASPT 172
 162 SWEGDANADYKQAIPLRNAGLNTLMDVDAAGQGPQSIHGYGREVENADPQRTMS 221
 173 CVLGRGSS--EYITFAIRSIDPDGVIVGSPWMSQ-----DHLADAPVSHSNMVA 224
 222 IHMTYAGNAGSOVRINIDRYLNDLALVIGEGHRTNGVDDEATMSISEQRGVQMLA 281
 225 LHF--YSGTGFQFLRIRITVAMNGAIFVTEWGTSDASGN----- 283
 282 WSMKGNPMEVYLDLSNDW--AGNNLTANGNTLVNGPYGLRETSRLSTVFGGSDGGT 338
 264 -----GGP-----YLPQSKEMIDFLNARKISWVN-----WSLADVEVESALMGASPTGA 309
 339 SPTLLYFEGSMQGTSGSLSGGFWATVTEWSSKGSLSL-KADTQLSSNSQHYLAVTQNTS 397
 310 GP-----MPNCRMKSGSSRSN-----PASNMRRGQNPAPAPATLSANGN-----AQVS 355
 398 LQONSRLQA--TYKHAN-----WGSYNGMTALRYKTG--HEGYTW-----SGSPVPI 442
 356 LTMNAVSGATSYTKRAITSSGSPYTNDRGVTAISYNTGLTNGTYYVYVVRASNS----- 411

QY 443 NGSSGTTLSLDLSNVONTLSQVREICVQFO-----SASDSSGQTSIVTDN 486
 DB 412 AGGSANSAQASATPAGGASTGNLVVQYKVGDTSAVDNMRKPSFNKN 459

RESULT 10

US-08-276-213-3

Sequence 3, Application US/08276213
 Patent No. 5536655

GENERAL INFORMATION:

APPLICANT: Thomas, Steven

APPLICANT: Laymon, Robert

APPLICANT: Himmel, Michael

TITLE OF INVENTION: GENE ENCODING FOR THE EI ENDOGLUCANASE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: National Renewable Energy Laboratory

STREET: 1617 Cole Boulevard

CITY: Golden

STATE: CO

COUNTRY: USA

ZIP: 80401-3393

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/276,213

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: O'Connor, Edna

REGISTRATION NUMBER: 29,252

REFERENCE/DOCKET NUMBER: NREL IR# 94-08

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303)231-1000

TELEFAX: (303)231-1098

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 521 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

US-08-276-213-3

Query Match 5.5%; Score 143; DB 1; Length 521;
 Best Local Similarity 21.0%; Pred. No. 0.00025;
 Matches 124; Conservative 70; Mismatches 180; Indels 216; Gaps 36;

32 ANSG-YVSGTLYDANGNPFVNGIN-----HGAWTKDQATTAIEGIANGA 79
 1 AGGGYWTSGREILDANNVPVRIAGINWGFETCNVYVHG-LMSRDY-RSMIDQIKSLGY 58
 80 NTVRIYLSGGQWTD--HTVRNLISLADNHLVAVPEVDATGYDSIASLNRADY 136
 59 NTRILPYS-----DILKPGTMSINFTQMN-----QDLQGLTSLQVMDKIYAY 103
 137 -----WTENRSALI-----GKEDVTIINIANEW 159
 104 AGGIGRIITLDNRDPCSGQSLMTYSSVSATWTISDQLAQLARKGNPTVVGFDLHNE- 162
 160 FGSWEGDANADG-----YKQAIPLRNA--GLNHTLV-----DAAGW----- 195
 163 --PHDPACWCGGCPIDMHLAERAGNAVLVSNPMLLFEVQVQSYNGDSYWGGLQCA 220
 196 GGP-----QSINHGYREVENADPQRTMPSIHMYEYAGNAGAV-RVINIDRYLN 244

Db 221 GQYFVYLVNPNRLYASADYATSY---PQ--TWSDPTFP---NNMGINKNNGYLEFN 272
QY 245 QDLALV-IGEGFH--RHTNGVDENTINSY---SEORGVG---WIAWSKNGPMEYILD 295
Db 273 ONIAVWMLGEFTLTOSTDTQTLVYLRPTAQYGADSFQWTFWMSNPDGDTGCI- 331
QY 296 LSNMAGNNLRAMKNTYNGPYGLAETSRSLVYTFGGSGDGTSP--TLYDFESSMGW 353
Db 332 LKDDW-----QTVDPV-----KDYLAIDKSIIDPVA----- 360
QY 354 TGSLSGGPMVATWESKSHLKADIQLSNSCH-----Y 389
Db 361 -----SASP-----SSQPSBSVSPSPSPASRTPPTPTPTASPTPTLTPTPTPT 409
QY 390 LHVITQNTSLQONSRIQATVK--HANNGSVNGMTAFLVYKTHG---YTW-----YSG-- 437
Db 410 ASPSPPTAASGACTASYQVNSDW---GNGFTVTVAV--TNSGSVATKTVTWSMTFGNQ 465
QY 438 -----SFYPLNGSSGTTSLDLSNQNLSQVREIGVQFQSASDSGOT 480
Db 466 TITSMNAAYVQNGQSVTARNKSYNNVIOQGQNTTFGFQASYSTGSAAPT 515

RESULT 11
US-08-604-913B-11
Sequence 11, Application US/08604913B

Patent No. 5712142
GENERAL INFORMATION:
APPLICANT: Adney, William S.
APPLICANT: Thomas, Steven R.
APPLICANT: Himmel, Michael E.
APPLICANT: Baker, John O.
APPLICANT: Chou, Yat-Chen
TITLE OF INVENTION: METHOD FOR INCREASING
TITLE OF INVENTION: THERMOSTABILITY IN CELLULOSE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: U.S.A.
ZIP: 80401-3393

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II (DOS) text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,913B
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,213
FILING DATE: 15-07-0-1994
ATTORNEY/AGENT INFORMATION:
NAME: Edna M. O'Connor
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: 95-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/384-7573
TELEFAX: 303/384-7459

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FEATURE:
NAME/KEY: E1-CAT
US-08-604-913B-11

Query Match 5.48; Score 142; DB 1; Length 358;
Best Local Similarity 23.18; Pred. No. 0.00017;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

QY 32 ANSGF-VYSCTIVDANPNFVWGIN-----HGNAWYKDAQTLAIEGANTGA 79
Db 1 AGGGTWTSGREILDANNVPRVINGINMFGEETCNVYHG--LMSHDY-RSMIDQIKSLGY 58
QY 80 NTVRYVLSDDGGQWTKDDY---HTVANTLSAEDNHLVAVPEVHDATGYDSIASLNRADV 136
Db 59 NTRIRPYS-----DLIRKPTMNSINFYQMN-----QDLQGLTSLQYMDKIYAY 103
QY 137 -----WTEMSALI-----GKEDVIYINIANEW 159
Db 104 AGQIGRLIILRHPRDCSGSALMTSSVSSEATWISDQLAALQRYKGNPTVVGFDLNE- 162
QY 160 FGSWEGDAMADG-----KKAIPRLRNA--GLNHTLV-----DAAGW----- 195
Db 163 --PHDPACWGGGDSIDWRLAERAGNAVLSVNPVLLIEVGVSQSYNGDSYWGNGLOGA 220
QY 196 GQFP-----QSIDYGREYFNADPQRTWESIMYEYAGGNASQV--RTNIDRYLN 244
Db 221 GQYFVYLVNPNRLYASADYATSY---PQ--TWSDPTFP---NNMGINKNNGYLEFN 272
QY 245 QDLALV-IGEGFH--RHTNGVDENTINSY---SEORGVG---WIAWSKNG----- 288
Db 273 ONIAVWMLGEFTLTOSTDTQTLVYLRPTAQYGADSFQWTFWMSNPDGDTGCI 332
QY 289 -PEWEYLDLSD 299
Db 333 KDDWQTVDPVD 344

RESULT 12
US-08-038-682-2

Sequence 2, Application US/08038682
Patent No. 5549897

GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: BERSKSTESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

Query Match 4.98; Score 129; DB 1; Length 1536;
Best Local Similarity 21.18; Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;

```
QY 1 LNNGFKIFSTTSLILASSILFVSGTSTANANG-----FYVSGTLLYDANGNPEV 52
DB 881 INSG-----NLTAGNTVNTAGNLTVESNANFAITNFTNNG--LFDKKGNSNI 929
QY 53 MRGINHAWYKQDQATTAIEGIANTGANTVRIYVSDGGQTKDDIHVRNLISLAEDNHL 112
DB 930 --STAKGARRKIDNSKNLSITNSSSTYRTIISGNTNNKNGD---NITNGSDTEM 983
QY 113 VAVPEVDATGYDSIAS--LNRAVDYIEMKSAIIGEDYIINIANEWFSGWEGDAMAD 170
DB 984 QIGDVSOKEGNLTISDKIN-----ITKQITKAGVGE-----NSDSD 1023
QY 171 GYKQAIPLRLAAGNLHTLMDVDAAGWGFPOSIDHGREVENADPQRTMFSIHMYEYAGG 230
DB 1024 ATNNANLTIKREKLQDNLISGFNK-----AETAKDSGLTIGTNSAD--GT 1072
QY 231 NASQVRNIDRYLQDNLALVIGEGHRTNGVDDEATIMSYSEORGWGLAMSKNGGPE 290
DB 1073 NAKKV-----TFNQVKSKISADGKH-----VTLHSKVEYSG-----SNNMTE 1110
QY 291 WEYLDLSDMAG-----NMLAMGNTIYNGPYGLRETSRLSTVFG-----331
DB 1111 ---DSSDNNAGLTIDAKNNTVNNNTSHKAVSISATSGEITTKGTTINATGVEITA 1166
QY 332 -----GSGDGTSPPTLYDFEGSMQGTGSSLSGCPNAVTEMS-----KGSHS 375
DB 1167 QTGSLIGIESSSGVTLTATGAL--AVSNISGNTVTVANSALTLTAGSTIKGES 1223
QY 376 LKADIQLS-----SNSOYHLVIONTSLOQNSRIQATVKAN---WGSVNGMTARLY 425
DB 1224 VTTSSOSGDIGTISGVEKATESLTQNSKIKATGGEANVTASATGTTGGTISGNTV 1283
QY 426 VATGCGIYWGSSFPVINGSSG---TTLSDLSNVONLSQVREIGVQFOSASDSGOTS 481
DB 1284 NVTANAAGDLYVNGAELNATGEGATLTTSKGKLTTEASSHTTSANGQVNLSDGSVAGS 1343
QY 482 IYIDNV 487
DB 1344 INAANY 1349
```

RESULT 13
US-08-302-832-2
Sequence 2, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Berenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: Of No. 5603938-typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 4.98; Score 129; DB 1; Length 1536;
Best Local Similarity 21.18; Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;

```
QY 1 LNNGFKIFSTTSLILASSILFVSGTSTANANG-----FYVSGTLLYDANGNPEV 52
DB 881 INSG-----NLTAGNTVNTAGNLTVESNANFAITNFTNNG--LFDKKGNSNI 929
QY 53 MRGINHAWYKQDQATTAIEGIANTGANTVRIYVSDGGQTKDDIHVRNLISLAEDNHL 112
DB 930 --STAKGARRKIDNSKNLSITNSSSTYRTIISGNTNNKNGD---NITNGSDTEM 983
QY 113 VAVPEVDATGYDSIAS--LNRAVDYIEMKSAIIGEDYIINIANEWFSGWEGDAMAD 170
DB 984 QIGDVSOKEGNLTISDKIN-----ITKQITKAGVGE-----NSDSD 1023
QY 171 GYKQAIPLRLAAGNLHTLMDVDAAGWGFPOSIDHGREVENADPQRTMFSIHMYEYAGG 230
DB 1024 ATNNANLTIKREKLQDNLISGFNK-----AETAKDSGLTIGTNSAD--GT 1072
QY 231 NASQVRNIDRYLQDNLALVIGEGHRTNGVDDEATIMSYSEORGWGLAMSKNGGPE 290
DB 1073 NAKKV-----TFNQVKSKISADGKH-----VTLHSKVEYSG-----SNNMTE 1110
QY 291 WEYLDLSDMAG-----NMLAMGNTIYNGPYGLRETSRLSTVFG-----331
DB 1111 ---DSSDNNAGLTIDAKNNTVNNNTSHKAVSISATSGEITTKGTTINATGVEITA 1166
QY 332 -----GSGDGTSPPTLYDFEGSMQGTGSSLSGCPNAVTEMS-----KGSHS 375
DB 1167 QTGSLIGIESSSGVTLTATGAL--AVSNISGNTVTVANSALTLTAGSTIKGES 1223
QY 376 LKADIQLS-----SNSOYHLVIONTSLOQNSRIQATVKAN---WGSVNGMTARLY 425
DB 1224 VTTSSOSGDIGTISGVEKATESLTQNSKIKATGGEANVTASATGTTGGTISGNTV 1283
QY 426 VATGCGIYWGSSFPVINGSSG---TTLSDLSNVONLSQVREIGVQFOSASDSGOTS 481
DB 1284 NVTANAAGDLYVNGAELNATGEGATLTTSKGKLTTEASSHTTSANGQVNLSDGSVAGS 1343
QY 482 IYIDNV 487
DB 1344 INAANY 1349
```

RESULT 14


```
Db 881 INSG-----NLTAGNIVNIAGNLTVESNANFKAITNFTFNVG--LFDNKGNSNI 929
QY 53 MRGINHGHWKQDQATTAIEGIANGTANTVRIVLSDGQMTKDDIHTVRNLISLAEDNHL 112
Db 930 --STAKGARPKDIDNSKNISITNSSSTYRTISGNTFNKNGDL---NITNGSDTEM 983
QY 113 VAVPEVHATGIDSLAS--LRAVDYNTLEMSALIGKEDVYIINIANEFSGMEGDAMAD 170
Db 984 QIGGVVSQKEGNLTISDKIN-----ITKOITIKAGVDE-----NSDSD 1023
QY 171 GYKQAIPLRNAGLNIHTLMDVDAAGQFPQSIHDYGREVNADPQRTMFSIHETAYAG 230
Db 1024 ATNNANLTIKREKELKTODLINSFNK-----AETAKDSDLLIGNTNSAD--GT 1072
QY 231 NASQVFNIDRYLNODLALVIGERGHRTNGDVDEATIMSISEQRGVGNLAMSWKNGPE 290
Db 1073 NAKKY-----TFNQVKDSKISADGK-----VTLHSKVEVTSG-----SNNTE 1110
QY 291 WEYLDLNDWAG-----NNLTFAMGNTIVNGPYGLRFSRLSTVFTG----- 331
Db 1111 ----DSSDNAGLITDAKNVTYNNNITSHKANVISATSGEITTKGTINATGVETLA 1166
QY 332 -----GGSDDGTSPTTIYDFEGSMQGTGSSLSGCPNAVTEWS-----KGSHS 375
Db 1167 QTSILGIGIESSSGVTLTATEGAL---AVSNISGNTVTVANSALTTLAGSTIKTES 1223
QY 376 LKADIQLS-----SNSQHYLHVIONTSLOONSRIQATVKHAN---MGSVNGMTARLY 425
Db 1224 VTTSSQSGDIGTISGTVKATESLITQNSKIKATIGEANVTSAITGIGTISGNTV 1283
QY 426 VRTGHGYTWYSGSFVPINGS--TTLISDLISNQNLSQVREIGVQFQASDSSGQTS 481
Db 1284 NVTANNGDLTVGNGAEINNTGGAATLTSSGKLTFASSHITSAKGOVNLSDQDSVAGS 1343
QY 482 IYIDNV 487
Db 1344 INANVY 1349
```

Search completed: December 19, 2000, 16:06:50
Job time: 392 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:00:19 ; Search time 245.85 Seconds

(without alignments)
126.482 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNGFKRISITSLIAS.....QSASDSGRTITDNYIVE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1614	61.9	516	2 JE0134	mannan endo-1,4-be
2	761.5	29.2	384	2 T37232	secreted beta-mann
3	731.5	28.1	363	2 S30386	mannan endo-1,4-be
4	638	24.5	1331	2 A48954	mannan endo-1,4-be
5	427	16.4	513	2 A37219	mannan endo-1,4-be
6	239.5	9.2	557	2 UC5487	cellulase (EC 3.2.
7	234	9.0	448	2 A27631	cellulase (EC 3.2.
8	226.5	8.7	409	2 A25156	cellulase (EC 3.2.
9	226	8.7	409	2 B25156	cellulase (EC 3.2.
10	220.5	8.5	505	2 S39962	endo-glucanase - Er
11	207	7.9	504	2 S54744	cellulase (EC 3.2.
12	203.5	7.8	570	2 S56132	cellulase (EC 3.2.
13	193	7.4	499	2 A27198	cellulase (EC 3.2.
14	188.5	7.2	508	2 A26874	cellulase (EC 3.2.
15	181	6.9	508	2 G69593	cellulase (EC 3.2.
16	180	6.9	486	2 I40548	cellulase (EC 3.2.
17	179	6.9	466	2 C42360	cellulase (EC 3.2.
18	173	6.6	783	2 UC5467	cellulase (EC 3.2.
19	171	6.6	499	2 J01111	cellulase (EC 3.2.
20	170.5	6.5	584	2 J01229	cellulase (EC 3.2.
21	169	6.5	822	2 J01611	cellulase (EC 3.2.
22	168	6.4	800	2 A29003	cellulase (EC 3.2.
23	166	6.4	941	2 S29043	cellulase (EC 3.2.
24	164.5	6.3	825	2 J50174	cellulase (EC 3.2.
25	160.5	6.2	510	2 T37341	probable glycolipid
26	156	6.0	611	2 J01177	endo-glucanase V (E
27	155	5.9	814	1 C2C1EM	cellulase (EC 3.2.
28	150	5.8	357	2 PC4404	cellulase (EC 3.2.
29	147.5	5.7	429	2 S29044	endo-glucanase A pr

ALIGNMENTS

Query Match	Best Local Similarity	Score	DB 2:	Length	516:
Matches 291; Conservative 84; Mismatches 87; Indels 28; Gaps 2;	61.9%	59.4%	Pred. No. 1e-102;		
RESULT 1					
JE0134					
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans					
N:Alternate names: endo-1,4-beta-mannanase					
C:Species: Bacillus circulans					
C>Date: 03-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999					
C:Accession: JE0134					
R:Yoshida, S.; Sako, Y.; Uchida, A.					
Biosci. Biotechnol. Biochem. 62, 514-520, 1998					
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene c					
A:Reference number: JE0134; M01D:9823274					
A:Accession: JE0134					
A:Molecule type: mRNA					
A:Residues: 1-516 <YOS>					
A:Cross-references: DDBJ:AB007123					
A>Note: the authors translated the codon CAA for residue 259 as Asn, CAA for residue					
A>Note: the translation of the nucleotide sequence 1294-1365 is not given in this c					
C:Keywords: glycoprotein; glycosidase; hydrolase					
30 146.5 5.6 747 2 BA7093					cellulase (EC 3.2.
31 140.5 5.4 426 2 A42649					cellulase (EC 3.2.
32 140.5 5.2 1335 2 T17508					glycoprotein Vp260
33 135 5.4 438 2 A47702					glucan 1,3-beta-gl
34 133 5.1 1091 2 G64964					hypothetical prote
35 132.5 5.1 1608 2 A28182					hemolysin A - Ser
36 132 5.1 1300 2 T00317					probable serine pr
37 131 5.0 517 2 T40798					cellulase (EC 3.2.
38 130 5.0 515 2 S20493					endo-glucanase - C1
39 129 4.9 566 2 A40589					cellulase (EC 3.2.
40 129 4.9 1070 2 S75712					cellulase (EC 3.2.
41 129 4.9 1536 2 A43855					high-molecular-we
42 129 4.9 2020 2 C48399					ABC-type transport
43 128.5 4.9 890 2 A30481					bacteriocin BCN5 -
44 127 4.9 32 2 PC4278					guar gum-degrading
45 126.5 4.9 428 2 S03767					cellulase (EC 3.2.
QY 27 TSTANNSGFYSGTLLYDANGNPYMGINGHAWYKQATTAEGTANTGATVRIYL 86					
DB 29 TKAHRAASGFYSGTKLDDATGQPFYMGVNAHMYKQDSTALPAKATGANTIRIYL 88					
QY 87 SSGGQMTKDIDHTVNNLSLAEDNHLVAVPEVHADATGYDSIASLNRADVWTEMSALIG 146					
DB 89 ANGHWTLLDDYVTNNILLCENKLIATLEVDHATGSDSLDDNNAVYWGIRKALIG 148					
QY 147 KEDVYIINANWPGSWEDAWADGYKQAPRLNAGNHTLMVDAAGGQPPDSIHNG 206					
DB 149 KEDRYIINANWYGTWCGVANANGYKQAPRLNAGNHTLVDSAGGQPPDSVKNKG 208					
QY 207 REVFANADPOKNTMFSIHMYEYAGNAGQVTRIDRYLNDALVIGFEFRTNDVDVEA 266					
DB 209 TEVLNADPLAKNTVFSEIHMYEYAGNASTYKSIDVLAKNLALIIIGEGGNTINDVDDEA 268					
QY 267 TIKSYSEORGVGWLAMSWKNGPEWEYDLSNDMAGNNITFANGNTIVNGPYGLRSTRLS 326					
DB 269 TIKSYSEKGVGLAMSWKNGSSDLAYLDMTDMAGNSITSGNTIVNGSNGIKATSVLS 328					
QY 327 TYFTGGSGDGGISPT-----LYDEGSGMGQWSSSLSG 360					
DB 329 GTF--GGVTPYISSPTSTPTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 386					
QY 361 GFAVTEWSSKSHLKDADIDLSNSOHLVHTQNTSLQNSRIQATVKAHMGSSVNGNM 420					
DB 387 GPWVINEWKAIOGLTKADVLSQNSHTSLYITTSQNLGKSKSLATYKHAHMGWIGNGNI 446					

F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-53/Product: cellulase #status predicted <CAT>

Query Match 9.2%; Score 239.5; DB 2; Length 557;
Best Local Similarity 23.1%; Pred. No. 6,76-09;
Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

```

QY 2 NNGFKRISI-----TLISLLASLIFVSGTS-----TANAN-----SGEYV 38
DB 3 NNVKKRILISIVAGAMALVAVNINVALETTISNLGNANVKKPSVGKGLQLNKNGI-- 60
QY 39 SGTLLDANGNPFVANGIN-HGHANYKDOATTAIEGIANATGA-----NTRYIVL- 86
DB 61 --KTLCDKDNFIOLRGKSTHLOMF-----PGVYNNNAFALSDMNSNTRYILAMY 110
QY 87 -SDGGWTKDDI-HYVRNLIISLAEDNHLVAVPEVHDATGDSIASLNRAVDYIEMRSAL 144
DB 111 VAEQGYATNPVYKQFVINGININAIANDMYIVDMHMMNGDNQNAVYSGAQSEFNDISTL 170
QY 145 IKREDVILINANEMWEGSWEG---DA--WADGVKQAIIP---RLRNAGLNLHILVDAAGWG 196
DB 171 YPNKNKILIELCEPENGEGVTNDATGAQYKSYATPIYOLLRKGNENLLIYGNPMS 230
QY 197 OFPOSIDHYGREVFNADPQRTMFSIHMYETAGN-ASQVRNTIDR-----VLNODLA 248
DB 231 QRDRLAAD-----NFINDSMTKSYHF--YSGTNPISYDINRDVNAAGNVEYALNHGA 282
QY 249 LVYGEGRHNTNDVDDEATIMSYSEQRGVGWLAMSKWGPEMEYIDLSDND----- 299
DB 283 VFATEWGTSLANG---TTGPYL-AKADAMDF--LNGNNISMCNFSINKDEKAAALNS 335
QY 300 -----WAGNMLIANG-----NTIVNGPYGLRETSRLSTVETGG 332
DB 336 LTLSDGSKLMDNDELITISGOYVBARIKGAYATPVDPTNGPAPKFFSSGFMDPNDG 395
QY 333 GSDG---GTSPTLLDFEGSGMOGWTGSSLSGCPMAVTESSKGSLSK-----ADIOS 383
DB 396 TTQGFQVNPSPITAINVENANNALKISNL-----NSKGSNDLSEGNFANVARS 445
QY 384 SNS-QHLYHIONTSLOONSRIOATYKHAMGVSNGMTARLVYKGHGY-----TW 434
DB 446 ADIMGOSINITYGDKLMDYIAPPY-----NYSIAIPQSSHGNGNTRALRW 496
QY 435 YSGSFVPIINGSGTLLSDLSNVQNLISQVREIGVQFOSASDS-----SGQISIT 484
DB 497 TNNVVAQTDGYKATLITISNDSPNFTIA-----TDADSVYTNMLFVGSNDNISL 550
QY 485 DNV 487
DB 551 DNI 553

```

RESULT 7

A27631
cellulase (EC 3.2.1.4) precursor - Clostridium acetobutylicum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium acetobutylicum
C:Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.
Appl. Environ. Microbiol. 54, 1289-1292, 1988
A:Title: Structure of an endo-beta-1,4-glucanase gene from Clostridium acetobutylicum p2
A:Reference number: A27631; MUID:88268074
A:Accession: A27631
A:Molecule type: DNA
A:Residues: 1-448 <ZAP>
A:Cross-references: EMBL:M31311; NID:g144789; PIDN:AAA23230.1; PID:g144790
A:Note: The authors translated the codon GAG for residue 116 as Gly, GAA for residue 263
C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 9.0%; Score 234; DB 2; Length 448;
Best Local Similarity 22.6%; Pred. No. 1.2e-08;
Matches 115; Conservative 75; Mismatches 185; Indels 134; Gaps 24;

```

QY 3 NGKRFKFSITLISLLASLIF-VSGTSTANANG-----FVSGTLLDANGNPFVNR 54
DB 9 NFFKRFESFLIAYV---MLFTYLGINTYKAEATTSFEGQLKYVGSQCDNSNGRKIQLK 65
QY 55 GI-NHGHANY-----KDOATTAIEGIANATVRIVL--SPGGQWT--KDI 97
DB 66 GMSHGLQWTVNPNVNDMSKFLDK-----GCVVYIRAAVNTNKGYSINPSSOK 115
QY 98 HYVRNLIISLAEDNHLVAVPEVHDATGDSIASLNRAVDYIEMRSALIKREDVILINAN 157
DB 116 EKIKKIYQDAIDIMMYVIDMHLISDNPNPTYEQAKSFQEMAE- YGKYSVYIETCN 174
QY 158 EMWGSWDAMADGK---QAIPLRNAGLNLHILVDAAGWQFOSIDHYGREVFND 213
DB 175 EPNQ---GTNANDIKRYANTYIIPAIADPNIIIVGISTWSDVDIAD-----N 223
QY 214 PQR--NTMFSIHMYETAGNASQVRNTIDRYLNODLALVYGEGRHNTNDVDDEATIMSY 271
DB 224 PLRYSNIMYICHF--YAGTHQSLRDKINYAMSGIAIFTEGTSDASGN----- 272
QY 272 SEQRGVGWLAMSKWGPEMEYIDLSDND---AGNNLTANGWTIVNGPYGLRETSRLST 327
DB 273 -----GGP---YLDESQWVDFMASKNI-SWTN-----WALCDKSEASA 307
QY 328 VFTGGSDGSGTSPYLLDFEGSGMOGWTGSSLS-----GCPMAVTESSKGSLSK 377
DB 308 ALKSGSSTTG-----GWTDSQTLTGLFVKKISIG-----SWTTSQTS 346
QY 378 ADIOSNSQHYHLYONTSLOONSRIOATYKHAMGVSNGMTARLVYK---TGHGTYW 434
DB 347 PTFISQSTYSAQTVLITISDNDSDVHYTTDGTPTSSPVTYSPITISKTTVKAFTT 406
QY 435 YSGSFVPIINGSGTLLSDLSNVQNLISQV 463
DB 407 KTG-----MTDSITSAVYTTISNTDPYKQV 431

```

RESULT 8

A25156
cellulase (EC 3.2.1.4) 1 - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
R:Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic Bacillus sp. s
A:Reference number: A51825; MUID:87056924
A:Accession: A25156
A:Molecule type: DNA
A:Residues: 1-488 <FUR>
A:Cross-references: GB:M4781; GB:X53449; NID:g142659; PIDN:AAA22301.1; PID:g142660
A:Experimental source: strain N-4, plasmid pK1
A:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 8.7%; Score 226.5; DB 2; Length 488;
Best Local Similarity 23.1%; Pred. No. 4.3e-08;
Matches 98; Conservative 67; Mismatches 157; Indels 103; Gaps 22;

```

QY 6 KRIFSITLISLLASLIFVSGTSTAN-----ANSGFVSGTLLDANGNPFVNRGI-NH 58
DB 2 KRLTIFIVFTLLA--LLFVGNSTSANNGSVVQNGQLSIONCOLVNEHGDVQLKGMSSH 59

```

```

01 59 GAAWTKDQATTAIEGANTGANTFVILYSDGGQWTKDD-----1HT----- 99
Db 60 GLQWFG-----OFTNIDTSIKMLDDMGITVFPAWMTSGGTYEDPS 101
02 100 ----VRNLISLAEQNLHVAVPEVHADATGYDSIASLNRADVYIEKRSALIGKEDFVIINI 155
Db 102 VAEKTKAEVAEALIDGIIYIIDMHLISDMDPNLYKEEAKEFDEH-SALYGSYPRVYIEI 160
03 156 ANEWFG---SWEGDAMADGY-KQALPRLNAGLNTLTAMVDAAGWQFPOSIHGYREVFVN 211
Db 161 ANEPNGHNVRW--DSHIKRYAEVPIVIRANDPNNVIIYGTATWS---QDVHEADNOLD 215
04 212 ADPQNTNESHIMTYEYAGGNAQVFRNTIDRYLNOGLALVIGEFGRHNGD----VDERT 267
Db 216 -DP--NVMYAEHF--YAGTHGQLRNQVDYALSRCGALIPVSSWGTSAATGDSGVLEDEAQ 270
05 268 I-MSYSEQRGVWMLAWS-----WKNGPWEWEYLDLSNDVAGNLTJLAWGNTIYNG 315
Db 271 VWIDPMDEENLSSMAWMSLHNDESSAALLPGANP-----TGGTATLALSSGAFVBE- 322
06 316 PYGLAETSLSTVFVWGGGSD---GCTSPITLYDFEGSMQWGTGSSL-----SGEPWA 364
Db 323 --KIRSESASIPSPDPTPSDDPDGPEDDPTPSD-PGEYFAMPDPNOIYTNWEIYHNGQLMQ 379
07 365 VTEM 369
Db 380 AKWMT 384

```

```

RESULT      9
B25156
cellulase (EC 3.2.1.4) 2 - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: B25156
R:Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J:Biochem.J. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic Bacillus sp. stre
A:Reference number: A91835; MUID:87056924
A:Accession: B25156
A:Molecule type: DNA
A:Residues: 1-409 <FUK>
A:Cross-references: GB:M14729; NID:g142655; PIDN:AAA22299.1; PID:g142656
A:Experimental source: strain N-4, plasmid pMK2
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

```

```

Query Match      8.7%: Score 226; DB 2: Length 409;
Best Local Similarity 23.2%: Pred. No. 3.6e-08;
Matches 104; Conservative 62; Pident 147; Indels 136; Gaps 24,

QY 6 KRIEFTISLTLASSILFVSGTSTA-----NANGSFVSGTFLTYDANGNPFWKGI-N 57
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 2 KRITTFIVLLMTLA-LFTIGNTAADDSVYEHHGQLSINSGLYNVDGEFVQLKGMS 60

QY 58 HCHANYKKOQTATAGSIANTGVARIYLSGGQWTKD-----IHT----- 99
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 61 HOLQWTG-----QFVNYEEMKMLRDMDGITVFRAAYTSGGGYEDP 102

QY 100 -----VRNLISAEENHLVAPEVDHACGYDSLASLRADVMIERSLIKREDVIYN 154
   |::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 103 SYKEEVYKEAVEALIDLGIVTIIDMHLISNDNNITYKEEKDFDEN-SELYDYDPVYIE 161

QY 155 IANWFVG---WEGD--AWADGKOAIPLRLAAGLNHTLMVAAGWGCPQSIEDYGREY 209
   |||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 162 IANEPNGSDVTMDNQIKPYAE---EVLPIARRNDPNIIIV---GGTGSQDVH-HAADN 214

QY 210 FNADQRRTMESIHRTETVAGNASOVIRINIDRYLNODLALVTEGRGHRTNGD----VDE 265
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||

```

```

Db 215 QULTDP--NMKAFHF--YAGHGNQLRQDVYDALDQGNALFVSMGTSSEATGGGVFLDE 270
QY 266 ATI-MSYEGGQGVGNLANS-----WKGNGPMEYELDISNDWAGNNLTAMGNTIV 313
Db 271 AQWIDPFDERLISVANNSTLHKDESSALIMPQASP-----TGCTWEALSPSGTFV- 322
QY 314 NQPYGLRSTRSLTYFTGG-----SDGCT-----SPTLLYD 345
Db 323 -----RKIRIRSATTPPSDTPPSDDPGSEHPDPGEPPDTPPSDGDYDPAWDPNTIYT 376
QY 346 FE-----GSW---QGWGSSLSG---GPW 363
Db 377 DEIVYHNGOLQWAKWNTIONQEPGDPYGFW 405

```

```

RESULT      10      .
S39962      endoglucanase - Erythrina carotovora
C:Species:   Erythrina carotovora
C:Date:      13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39962
R:Cooper, V.J.C.; Salmond, G.P.C.
Mol. Gen. Genet. 241, 341-350, 1993
A:Title:      Molecular analysis of the major cellulase (CelV) of Erythrina carotovora: evid
A:Reference number: S39962, MUID:94067016
A:Accession:  S39962
A:Status:     preliminary
A:Molecule type: DNA
A:Residues:   1*505 <COO>
A:Cross-references: EMBL:X76000; NTD:9434941; PIDD:CAM53592.1; PID:9434942

```

[illegible]

RESULT 11

554744 cellulase (EC 3.2.1.4) CelV precursor - *Erythrina carotovora* (SCC 3193)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Erythrina carotovora*

A:Variety: SCC 3193

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S54744; S44996

R:Maee, A.; Heikkinen, R.; Palva, E.T.

Mol. Gen. Genet. 247, 17-26, 1995

A:Title: Structure and regulation of the *Erythrina carotovora* subspecies *carotovora* SCC3193

A:Reference number: S54744; MUID:95231512

A:Accession: S54744

A:Molecule type: DNA

A:Residues: 1-504 <MAE>

A:Cross-references: EMBL:X79241; NID:g493492; PIDN:CAA5823.1; PID:g493493

C:Genetics:

A:Gene: celV1

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-33/Domain: signal sequence #status predicted <SIG>

F:33-504/Product: cellulase #status predicted <MAT>

Query Match 7.9%; Score 207; DB 2; Length 504; Best Local Similarity 22.3%; Pred. No. 9.6e-07; Matches 110; Conservative 63; Mismatches 191; Indels 130; Gaps 22;

```

QY 6 KRIFSTLSLLASSTLFWSTGTSTANANG-FTVSGTLYDANGPFWRG-NGHMY 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 10 RKLLTGVTMTLMGMSFSFSLATPVETHGQSIENGRLVDGKRVQGLSSNGLOW 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 KQATTAIGTANTGANTVRYLSDGQWTKD--IHTFNLISLAEDNHLVAPEV--- 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 70 GD-----YVKKDSMKHLRDMGINTFRVAMTAE-NGYLANSLANK 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 -----HDATGYSIASLNRADVWTENKSLIGKEDVTYINIANEW 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 111 VKENVAAGLGYIIIDHTLSDNDPNTYKAQAKIFFEEM-AGLGNPNVITYETANEP 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 FGS--WEGD--AMADGVKQALPRLNAGLNHTLVDAAGCGFPOSIDHYGREVFNAQD 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170 NGSVTMNGQIRPYA---LEVTDTIRSKDPNLIIVSGSTW---QDIHD--AADQOLD 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 RNTMFSIMTEYAGNAGQVRINIDRYLNQDLALVIGFGRHTND-----VDEATWS 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 221 PNTLYALHF--YAGTHGQFLRDRIDYQSGRAALFVSEMGTSASGNGGFLPEESQWID 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 YSEORGVGLAMSKWKGKPEWEYL-----DISNDMAGNNLTMANGTYNGYGLRETSRLS 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 279 FLNKRGISWNWMSLSDKSETSAALVAGASKSGGTEONLSISGFTV-----REQIR-- 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 327 TVTGGSDGDSPTTLYDFEGSMQGTGSLSGPWAATWESKSGSHLKADIOSNS 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 330 ---AGALSGDPTTMTPEPTNPNGTIG-----DIVL----- 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 387 QHLYHTVQNTSLQONSRIQATVKAHNGSVG--NGATPALLYTKTGCTYWSGSEFPIN 443
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 360 -QYRNVNPNPS---DDAIRMAFNKIKNGTSPKIKSLQVRYFHD-----D 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 444 GSSGTTLSDLSNV 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 402 GKPGANLEFVDWAV 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12

556132 cellulase (EC 3.2.1.4) precursor - *Pseudomonas fluorescens*

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Pseudomonas fluorescens*

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S56132; S56127; S54798

R:Hall, J.; Black, G.W.; Ferreira, L.M.A.; Millward-Sadler, S.J.; All, B.R.S.; Hazlewood

Biochem. J. 309, 749-756, 1995

A:Title: The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas*

A:Reference number: S56127; MUID:95366948

A:Accession: S56132

A:Molecule type: DNA

A:Residues: 1-570 <HAL>

A:Cross-references: EMBL:X86798; NID:g806573; PIDN:CAA60493.1; PID:g806574

A:Accession: S56127

A:Molecule type: protein

A:Residues: 40-54 <HAM>

C:Genetics:

A:Gene: celE

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: bacterial cellulase-binding domain homology; glycosidase GWCW domain h

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-33/Domain: signal sequence #status predicted <SIG>

F:40-570/Product: cellulase #status experimental <MAT>

F:380-417/Domain: glycosidase GWCW domain homology <GWC>

F:472-569/Domain: bacterial cellulase-binding domain homology <BCB>

Query Match 7.8%; Score 203.5; DB 2; Length 570; Best Local Similarity 23.3%; Pred. No. 2e-06; Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

```

QY 14 SLTLASSILFVSGTSTANANGFTVSGTLYDANGPFWRGIN-----HGHWYKD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 22 AIIILGSGILGCGVSAQADVAPLSVOGNKIL-ANGOPASESGSLFWSNTEWGEKYNA 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 QATTAIGTANTGANTVRYL---SDGQWTKD---KDHTFNLISLAEDNHLVAPEV 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 81 QVWSWLR--SWNNKTLVRAAGVEDEGGYITDPAANDRYQVVD-AIANDMYIIDMS 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 HDATGYSIASLNRADVWTENKSLIGKEDVTYINIANEWFG-SWEGMADGYQA-1 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 138 HNAHQYOS-----QALFQDEM-ARKYGANNHYEYENEPLOYWSMT--IKPYQAVI 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 PRLNAGNHTLVDAAGWGFPOSIDHYGREVFNAEP--QNTMFSIMTEYAGNAGS 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 190 AAIRAIDPDNLIIVGTPWSQ-----DYVAANDPITGYQIVATTLF--YAGTHGQ 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 QVRETRIDVNLQDIALVIGFGRHTNDYDEATIMSYSEORGVGLAMSKNGPEWEY 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 240 YLRKAQIALNRGIALVYTEMGSYVNAAGD---GAVANSE----- 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 294 IDLSNDMAGNNLTMANGTYNGYGLRETSRLSTVFTGGSGDGTSPITLYDFEGSMQGW 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 276 ---TNAWVSFMKT---NHISNANMALNDKYEGASALVPGASANG-----GW 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 354 TGSSSL-SGGPNA---VTEW-----SSKGSLSLKADIOSNSQHLHYIONTSLQONSRIQ 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 316 VNSQLTASGALAKSLISCPMTNTSSSSSAVSSQTVSSSQ--APVSSSSSTASSAVS 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 406 ATV--KHANWGSVNGMGTARLYTKTGTY-----YGSFVPINGSGGTITSL 452
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 374 SAVSQGQCM-----YGLIYPLCSTTTNGMENNASCARATCGQAPMGIVGGSTSS- 428
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 453 DLSNVQNTSQRVEIGVQFOSASDSSGQTSI 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 429 -----QASSSVRSSSSSIYSSRSSSSSV 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13

A27198 cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain IF03034)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Bacillus subtilis*

C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999

C:Accession: A27198

R:Nakamura, A.; Uozumi, T.; Bepko, T.
 Eur. J. Biochem. 164, 317-320, 1987
 A:Title: Nucleotide sequence of a cellulase gene of *Bacillus subtilis*.
 A:Reference number: A27198; MUID:87190397
 A:Accession: A27198
 A:Molecule type: DNA
 A:Residues: 1-499 <NAB>
 A:Cross-references: GB:M28332; NID:g142670; PIDN:AA22307.1; PID:g142671
 A:Experimental source: strain IFO3034
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 8.5e-06;
 Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

```

QY 6 KRFSTTSLASSLTF-----VSGTSTANNSG-FVSGTTLTDANGNPFVNRG 55
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 KRSISFICLLIYVLMGGLQSPASAGTKPAKNGLSIKGTOLVNRGKAVQLKG 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 I-NHGAWYKQDQATTAIEGIANGTAVRIYLSDGQWTKD-----IHT----- 99
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 ISSHGLQWTD-----FVNKDSLKLRLDMGIVTFRPAAYTTADGGYI 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 100 -----VRNLSLAEDNHLVAPEVDATGYDSIASINRAVDYIEKRSALIGKEDTV 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 104 DNPVKNKKEAVEAKEIGIYIIDWHLNDGNPNQHEKAKDFEKEK-SSLYGNTPNV 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 IINIANEMFG--SWEGD--AMADGYKQAIPLRNAGLNTLWDAAGCOFPOSIDHYR 207
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 163 IYEIANEPGDVNWKMDIPYAE--EVSIVIRKNDPDITIV--GIGTWSDQVND--- 213
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 208 EYFNADPQ---RNTMSIMYEYAGGNASQVNTNIDRYLNODLALYIGERHRTNGD--- 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 214 ---AADQDLKANVMYALHF--YAGHGOSLRDKANVYALSKAPIFVTWGTSDASGNG 268
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 263 ---VDEATISYSEORGVGLWMSKNGNPEWY--EYLDLSN-DNAGNNTLWAGNTIYNG 318
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 VFLDQSR-----EWLNTYDSKRIISWVNWTL----- 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 319 LRETSLSLSTFTGGSGDGTSTLYLDFEGSMQGTSSLSGGPMAVTEWMSKSHSLKA 378
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 294 -----SDKQESSALK-----PGASKTG-WEPLTDLTASGTF-VRE 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 379 DI--QLSSNSOHYLAHYQNTSLQNSRIQATYKHAMGSVGNMTRLYVTKGHTY--- 433
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 328 NLGNKDSSTKEPPEPAQNPQAEENG-ISOYKKAAGGAGVNSQNIQPIQLIKNGNATVDL 386
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 434 -----WTSGSFVPISSGTTSLDLSNV 457
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 387 KDVTARYWYNAR-----NKGQNECDYAGI 411
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14
 A26874
 cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain DLG)
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
 C:Accession: A26874; B26874
 R:Robson, L.M.; Chambliss, G.H.
 J. Bacteriol. 169, 2017-2025, 1987
 A:Title: Endo-beta-1,4-glucanase gene of *Bacillus subtilis* DLG.
 A:Reference number: A26874; MUID:87194581
 A:Accession: A26874
 A:Molecule type: DNA
 A:Residues: 1-508 <ROB1>
 A:Cross-references: GB:M16185; NID:g143007; PIDN:AA22496.1; PID:g143008
 A:Experimental source: strain DLG

A:Accession: B26874
 A:Molecule type: protein
 A:Residues: 39-53 <ROB2>
 A:Experimental source: strain DLG
 A:Note: The authors believe Met-1 and Met-2 may be alternate initiators
 C:Comment: The low molecular weight of the mature protein suggests carboxy-terminal
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: alternate initiators; extracellular protein; glycosidase; hydrolase;
 F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 7.2%; Score 188.5; DB 2; Length 508;
 Best Local Similarity 21.1%; Pred. No. 1.8e-05;
 Matches 78; Conservative 75; Mismatches 103; Indels 83; Gaps 17;

```

QY 6 KRFSTTSLASSLTF-----VSGTSTANNSG-FVSGTTLTDANGNPFVNRG 55
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11 KRSISFICLLIYVLMGGLPSPASAGTKPAKNGLSIKGTOLVNRGKAVQLKG 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 I-NHGAWYKQDQATTAIEGIANGTAVRIYLSDGQWTKD-----IHT----- 99
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 71 ISSHGLQWTD-----FVNKDSLKLRLDMGIVTFRPAAYTTADGGYI 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 100 -----VRNLSLAEDNHLVAPEVDATGYDSIASINRAVDYIEKRSALIGKEDTV 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 113 DNPVKNKKEAVEAKEIGIYIIDWHLNDGNPNQHEKAKDFEKEK-SSLYGNTPNV 171
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 IINIANEMFG--SWEGD--AMADGYKQAIPLRNAGLNTLWDAAGCOFPOSIDHYR 207
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 172 IYEIANEPGDVNWKMDIPYAE--EVSIVIRKNDPDITIV--GIGTWSDQVND--- 222
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 208 EYFNADPQ---RNTMSIMYEYAGGNASQVNTNIDRYLNODLALYIGERHRTNGD--- 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 223 ---AADQDLKANVMYALHF--YAGHGOSLRDKANVYALSKAPIFVTWGTSDASGNG 277
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 263 --VDEA-TTMSYSEORGVGLWMSKNGPEWEY---DLSNDNAGNNTLWAGNTIYNG 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 278 VFLDQSRWLNLYDSKNISWVNWNLSDKQESSALKPGASKTGWPLDITLTSAGTFVREN 337
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 316 PYGLRETSR 324
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 338 INGTDSIK 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 15
 G69593
 cellulase (EC 3.2.1.4) bglC precursor - *Bacillus subtilis*
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: G69593; A26114; I40353; S49103; I39803
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Enliant, K.D.; Erlington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lartino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadleir, J.; Sato, T.; Scani
 A:Authors: Schleich, S.; Schwoerer, R.; Scoffone, F.; Sekiyuchi, J.; Sekowska, A.; Se
 akouchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchida
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yate, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: G69593
 A:Molecule type: DNA
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Residues: 1-508 <KUN>

A:Cross-references: GB:Z99113; GB:AL009126; NID:92634090; PIDN:CA813696.1; PID:el18347.1;
 A:Experimental source: strain 108
 R:Macay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; Sel
 Nucleic Acids Res. 14, 9159-9170, 1986
 A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
 A:Reference number: A26114; MID:87066783
 A:Accession: A26114
 A:Molecule type: DNA
 A:Residues: 10-508 <MAC>
 A:Experimental source: strain PAP115
 A:Note: Part of this sequence, including the amino end of the mature form, was confirmed
 R:Landahl, V.; Aa, K.; Tronsmo, A.
 Antonie Van Leeuwenhoek 66, 327-332, 1994
 A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C
 A:Reference number: I40353; MID:95225656
 A:Accession: I40353
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-291, 'N', 293-508 <LIN1>
 R:Landahl, V.; Aa, K.
 Submitted to the EMBL Data Library, June 1992
 A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
 A:Reference number: S24239
 A:Accession: S24239
 A:Molecule type: DNA
 A:Residues: 10-291, 'N', 293-508 <LIN2>
 A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
 A:Experimental source: strain CK-2
 R:Wolf, M.; Geetzl, A.; Borriess, R.
 Submitted to the EMBL Data Library, December 1993
 A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons
 A:Reference number: S49103
 A:Accession: S49103
 A:Molecule type: DNA
 A:Residues: 10-508 <WOL>
 A:Cross-references: EMBL:Z29076; NID:g509266; PIDN:CA82317.1; PID:g509267
 R:Seo, Y.S.; Lee, Y.H.; Park, U.H.; Kang, H.
 Korean J. Microbiol. 24, 236-242, 1986
 A:Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis
 A:Reference number: I36003
 A:Accession: I36003
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'P', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36-
 A:Cross-references: GB:M38634; NID:g142657; PIDN:AAA22300.1; PID:g142658
 A:Experimental source: strain ATCC 6633
 C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
 C:Genetics:
 A:Gene: bglc
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 C:Keywords: cellulose degradation
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:39-508/Product: cellulase #status predicted <MAT>

Query Match 6.9%; Score 181; DB 2; Length 508;
 Best Local Similarity 19.3%; Pred. NO. 5.8e-05;
 Matches 100; Conservative 102; Mismatches 186; Indels 128; Gaps 26;

QY 6 KRFFSTLLSLLASTLF-----VSGTSTANANSG-FYVSGTTLTDANGNPFVARG 55
 DB 11 KNSISFICLLTLTLMGMAISPAASAGITPAKKNQGLSIRKQGLVNRDGAVALQKG 70
 QY 56 I-NHGAWYKDOAT-TAIEGIANTGANV---RIVISDGGQWTKDDI-HTVANKLSLAE 109
 DB 71 ISSHGLQWGEYVKNKSLKMLRDWGIIVFRAMAYTDAGYIDNPSVKNKKEAVEAFAKE 130
 QY 110 NHELVANPEVHDANGYSIASLNRAVDYWTIEMSGALGKEDTVIINANEMFG--SMEGD- 166
 DB 131 LGIYIYIHWIILNDGNPNONKREKAEFEKEM-SSLGNTPNVYIETLANEPNDGVNKKRDI 189

QY 167 -ANADGYKAIPLRLNAGLNHTLAMPDAGWGQFPOSIDYGREVENADPO---RTIMESI 222
 DB 190 KPYAE---EIVSIVIRKNDNDNIIIV---GTGTWSDVDND-----AADDOLKDNAMVMTAL 237
 QY 223 HMYEYGGNAGQVRTNIDRVLNODLALVIGFGRHTNGD---VDEA-TIMSYEQRCV 277
 DB 238 HF--YAGTHQFLADKANALSKGAPITFTEWGTSDASGNGSVFLDQSRKMLKIDSXTI 295
 QY 278 GTLAWSKNGKPEWEYL---DISNDMAGNNLTAMGNTIVNCPYGLRERSR----- 324
 DB 296 SWVNMVNSDKQESSSALKPKPASKTGWRSLDSASTFRENILGKIDTKDIPETPSKD 355
 QY 325 -----LSVPELGGSDGDTPTTLTYDEGSMQGTGSSLSGGPWAATWESSKGSLSIK 377
 DB 356 KPTQENGISVQYRAG-----DGSW-----NSNQIR 380
 QY 378 ADIQLSSN-----SGHYLHVIONTSLOQNSRIQATVKHAMNGSVNGMTARLY- 426
 DB 381 PQLQIKKNGNTYVDLKNVYARWYKAKKNG--QNF-----DCDYAQIGCGNVTIKFVY 431
 QY 427 --KTGHGTYW-----SGSFVPINGSSGTTLSLDSLN 456
 DB 432 LHKPQOGADTYLELGFKNGLTAP--GASTGNTQLRLHN 467

Search completed: December 19, 2000, 16:10:59
 Job time: 640 sec

This Page Blank (usp10)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:29:24 ; Search time 131.22 Seconds
(without alignments)
119.300 Million cell updates/sec

Title: US-09-339-159-2
Perfect score: 2607
Sequence: 1 LNNGFKRISITSLILASS.....QSASDSGGOTSTIDNVIYE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues
Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	737.5	28.3	383	1	MANA_STRLI
2	638	24.5	1331	1	MANB_CALSA
3	427	16.4	513	1	MANB_BACSM
4	234	9.0	448	1	GUN_CLOAB
5	229.5	8.8	400	1	GUNS_BACAG
6	226.5	8.7	488	1	GUN1_BACSA
7	226	8.7	409	1	GUN2_BACSA
8	220.5	8.5	505	1	GUNV_ERMCA
9	219.5	8.4	444	1	GUNV_ERMCA
10	207	7.9	504	1	GUNV_ERMCA
11	188.5	7.2	499	1	GUN1_BACSU
12	181	6.9	499	1	GUN2_BACSU
13	180	6.9	466	1	GUNS_THERU
14	171	6.6	499	1	GUN3_BACSU
15	170.5	6.5	584	1	GUN3_CLOAE
16	168	6.4	800	1	GUN_BACSI
17	166	6.4	941	1	GUNA_BACSE
18	165.5	6.3	459	1	GUNA_STRLI
19	164.5	6.3	825	1	GUN3_BACSA
20	160.5	6.2	510	1	YXPD_SCHRO
21	155	5.9	814	1	GUNE_CLOFM
22	151.5	5.8	562	1	GUN1_ACICE
23	147.5	5.7	429	1	GUNA_BUTFI
24	146.5	5.6	747	1	GUND_CELFI
25	141	5.4	312	1	GUNA_RUMAL
26	140.5	5.4	426	1	GUN_BURSO
27	133	5.1	1039	1	AC43_ECOLI
28	132.5	5.1	1608	1	HLXA_SERVA
29	132	5.1	438	1	EXG_CANAL
30	131	5.0	515	1	GUND_CLOCL
31	131	5.0	517	1	GUNA_CLOLO
32	129	4.9	566	1	GUNG_CLOFM
33	129	4.9	2003	1	YDBA_ECOLI

34	128.5	4.9	890	1	BCN5_CLOAE
35	127	4.9	426	1	GUN2_ERMCH
36	123.5	4.7	645	1	LIP1_PHOUD
37	123.5	4.7	711	1	CDGT_BACST
38	122.5	4.7	1039	1	GUNB_CALSA
39	118.5	4.5	1829	1	FRPC_NEIMC
40	116.5	4.5	440	1	GUNB_CLOCL
41	116.5	4.5	976	1	PIBP_ADEB3
42	115	4.4	1656	1	OMP_RICUA
43	114.5	4.4	865	1	HYRE_ECOLI
44	114	4.4	802	1	PERC_SALTY
45	114	4.4	2334	1	WAPA_BACSU

ALIGNMENTS

RESULT 1
MANA_STRLI STANDARD; PRT; 383 AA.
ID MANA_STRLI
AC P51529;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MANAN ENDO-1,4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-MANNANASE) (1,4-BETA-D-MANNAN MANNANOHYDROLASE).
GN MANA.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Streptomyces; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 36-42.
RC STRAIN-66 / 1326;
RX MEDLINE; 93207541.
RA Arcand N., Kuepfel D., Paradis F.W., Morosoli R., Shareck F.;
RT "Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence of the manA gene and characterization of the enzyme.";
RL Biochem. J. 290:857-863(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN-66 / 1326;
RA Shareck F.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES CELSIUS AND A PH OF 6.8.
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; M92297; AAA26710.2; -;
DR INTERPRO; IPR001547; -;
DR PFM; PFM0150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
KW Hydrolyase; Glycosidase; Signal.
FT SIGNAL
FT CHAIN 1 35
FT DOMAIN 36 383
FT DOVAIN 336 340
SEQUENCE 383 AA; 39681 MW; SDB4B407C64E9AC3 CRC64;
MANAN ENDO-1,4-BETA-MANNOSIDASE.
POLY-GLY.
Query Match 28.3%; Score 737.5; DB 1; Length 383;
Best Local Similarity 44.4%; Pred. No. 1.8e-43;

Matches 160: Conservative 56; Mismatches 121; Indels 23; Gaps 9;

QY 13 LSLAASLLEFVSGTSTANAN---GFVYSGTTLTDANGNPPVAGINHGAWYDOATT 69
DB 19 LGLFA-----LAPSAGAAEAAAGIHVSNGRYVEGNSAFVAGVNAHYWYDR-TG 72
QY 70 ATEGIANCANTVRIYVLSGGGQKDDIHVNLISLAEDNHLVAVPEHATGY--PSI 127
DB 73 SIADIAAGANTVRYVLSGGGQKDDIHVNLISLAEDNHLVAVPEHATGY--PSI 132
QY 128 ASLNAVDYIEMRSAL-TGKEDYIINIANEMFSGWEGDAMADYKQAPLRLNAGIN 186
DB 133 TSLDQAGDYVGVKSAWMAEDYVYVNGINPEFGNTAAWTDATKSAIGLRGAGLGH 192
QY 187 TLMNDAAGWQ-FPOSIDYGREYVNAPORTNFSIHMEYAGNASVYRNIVRLNQ 245
DB 193 ALMVAAPMGQDWSGTMSNAVSASDPRATVSIHNG-VYDTAEVRDYLNAFVGN 251
QY 246 DLATVIGFGRHRTNGDVEATINSYSEQRYGMLAMSKWNGPMEYDLISNDVAGN 305
DB 252 GLPIYVGEFGDQSDNGDEDAIMATAGSLGVYLGKMSNGGVEYIDMNGDPNSL 311
QY 306 TAWGTIYNGPYGLRETSRLSTVF-TGGGSDGGSPT-----TLVDEGSMQWGTGS 356
DB 312 TSMGNRIYSGNGIATSRATATVYGGGSGTGTAIPNGYPCVNGGASDPDGDGMEWS 371

RESULT 2
MANB_CALSA STANDARD; PRG: 1331 AA.

AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BETA-MANNANASE/ENDOGUCANASE A PRECURSOR (INCLUDES: MANNAN ENDO-1,4-MANNANOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-MANNANASE); ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULOSE)).
GN MANA.
OS Caldocellum saccharolyticum (Caldocellulosirupor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
RN Thermomonas group; Caldacellulosirupor.
[1]
NM SEQUENCE FROM N.A.
RX MEDLINE: 93119139.
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme";
RL Appl. Environ. Microbiol. 58:3864-3867 (1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE: 91247819.
RA Luthi E., Jussant N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'";
RL Appl. Environ. Microbiol. 57:694-700 (1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC LINKAGES IN MANNANS, GALACTOMANNANS, AND GALACTOGLUCOMANNANS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.1sb-sib.ch/announce/> or send an email to license@1sb-sib.ch).

CC EMBL: L01257; AAA71887.1; -;
CC DR EMBL: M35063; AAA72861.1; -;
CC DR PIR: B43745; B43745.
CC DR PIR: A48954; A48954.
CC DR HSSP: Q06851; INBC.
CC DR INTERPRO: IPR001547; -;
CC DR INTERPRO: IPR001956; -;
CC DR PRAM: PR00942; CBD_3; 2.
CC DR PRAM: PR00150; cellulase, 1.
CC DR PROSITE: PS00659; GLYCOSYL-HYDROL-F5; 1.
CC DR Hydroxylase; Glycosylase; Cellulose degradation; signal;
CC K W Multifunctional enzyme.
CC FT SIGNAL 1 41
CC FT CHAIN 42 1331
CC FT DOMAIN 42 325
CC FT DOMAIN 326 361
CC FT DOMAIN 362 518
CC FT DOMAIN 519 564
CC FT DOMAIN 565 720
CC FT DOMAIN 721 780
CC FT DOMAIN 781 1331
CC FT ACT_SITE 162 162
CC FT ACT_SITE 257 257
CC FT CONFLICT 338 338
CC FT CONFLICT 340 346
CC FT SEQUENCE 1331 AA; 146892 MW; FFECA51BBD8D80E0 CRC64;

Query Match 24.5%; Score 638; DB 1; Length 1331;
Best Local Similarity 31.4%; Pred. No. 5.9e-36;
Matches 166; Conservative 84; Mismatches 208; Indels 70; Gaps 14;

QY 6 KRISITSLAASLLEFVSGTSTANANSGFYVSGTTLTDANGNPPVAGIN 57
DB 9 KRWLSVLTVEFLNLEIFIANVTILPRVGAATSN-----DGVKKIDTS-----TLIGTN 57
QY 58 HGAWYKDOATTAEGTANTGANTVRIYVLSGGGQKDDIHVNLISLAEDNHLVA-VP 116
DB 58 HAHCHYRRLRLTALRGISWGNVRYVLSNGYKWKIPASEVANIISLSLGFKATIL 117
QY 117 EVDHATGY--DSIASLRAVDYIEMRSALIGKEDYIINIANEMFSGWEGDAMADGYK 173
DB 118 EVDHATGY--DSIASLRAVDYIEMRSALIGKEDYIINIANEMFSGWEGDAMADGYK 177
QY 174 QAIRPLRAGINRLTAMDADAGWQ-FPOSIDYGREYVNAPORTNFSIHMEYAGGNA 232
DB 178 NAIKRLRAGKFKHIMVADAPMGQDWSMTMDNMSIEMDPLRLNVFSIMYG-VYNTA 236
QY 233 SOVRNIDRVNODLAVYIGFGRHRTNGDVEATINSYSEQRYGMLAMSKWNGPMEY 292
DB 237 SKVEEYIKSPDKGLPLYIGFGRHRTNGDVEATINSYSEQRYGMLAMSKWNGPMEY 296
QY 293 YLDSNDVAGNNTLAWGNTIYNGPYGLRETSRLSTVFETGGSDGDT-SPTLLYDEGSMQ 351
DB 297 YLDMVNNMPPNPPFWGQWYKTAIGTSSPTLPSTVPTPTPTPTPTATPTPTPTPT 356
QY 352 GWTGSSLSG-----GPAVYEMSKSGSLADQLQSLNSQHYLHV 393
DB 357 PVSPATSGQIKVLYANKETSTNTIRPMV-----KVNSSGSSIDLSVTLTRYWTV 410
QY 394 QNTSLQNSRQIATVYKHWGNSVGNWGTARLYTGR-----GYWWSGSPVPI 442
DB 411 -----DGRQSAI--SDMAQICASNTVTFYVLSVSSVSGADYLLTEGFSGAGQLPG 462
QY 443 NGSGGTLISLDSNVONLSQVREIG-VQFQASDSSGQTSIYDINVI 489
DB 463 KDTGEIQRFKDKDMVNTNGQNDMSWISQMSYSEKNTKVAIDGVLY 510


```

CC      1- SUBUNIT: MONOMER.
CC      -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC      HYDROLASES) :
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; AF067426; AAC19169.1; -
DR      PDB; 1A3H; 16-MAR-99.
DR      PDB; 2A3H; 16-MAR-99.
DR      PDB; 3A3H; 16-MAR-99.
DR      INTERPRO; IPR001547; -
DR      PFAM; PFC00150; cellulase; 1.
DR      PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR      Cellulose degradation: Hydrolase; GLYcosidase; 3D-structure.
KM      ACT_SITE 165; 165
FT      ACT_SITE 254 254
SQ      SEQUENCE 400 AA; 44702 MW; 3F9C6FB9BC36FEFF CRC64;
      NUCLEOPHILE
      PROMON DONOR.
      8.8%; Score 229.5; DB 1; Length 400;
      Best Local Similarity 23.4%; Pred. No. 9.5e-09;
      Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;
      6 KFIPTSLTLLSSILTFVSGISTANAN-----GFLVSGTLLTDANGNPFVWRI-NH 58
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
      2 KKTITFVILMTVALEFISGNTTAAANDSVVEHQGLSTISNGELVYENGEYQLKGNSSH 61
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
      59 GAAW-----KQATTAIEGINTGANTVRIVL-SDGQWTKDI-HIVRNL 103

```

[illegible]

RA Fukumori F., Sasahara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RL *Bacillus* sp. strain N-4 and their strong homology."; *Bacteriol.* 168:479-485(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOMYDOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14781; AAA22301.1; -.
 CC PIR: A25156; A25156.
 DR INTERPRO: IPR001547; -.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase.
 FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 488 AA; 54264 MW; 97248E45D28D3B5 CRC64;

Query Match 8.7%; Score 226.5; DB 1; Length 488;
 Best Local Similarity 23.1%; Pred. No. 2e-08;
 Matches 98; Conservative 67; Mismatches 157; Indels 103; Gaps 22;

QY 6 KRISITSLILASSILFVSGTSTAN-----ANSGFVSGTLLYDANGNPFVWRI-NH 58
 DB 2 KKITITFIVLMTLA-LFIVGNSTSNAGSVYEGNQLSIQNGVLNHEHDDPVOLKGMSSH 59
 QY 59 GHAWKQDATTALBGIANTGANTVRIYLSDGQWTKD-----IHT----- 99
 DB 60 GLQWYG-----QFVNYDSIKMLRDWMTITVFRAAMYTSSGGYIEDPS 101
 QY 100 -----VRLISLADNHLVAPEVEDATGYDSIASLNRAVDYIEMRSALIGKEDYIINI 155
 DB 102 VKREYKAEVAEALDGIYIIDWHLISNDPNRIYKEAKEPFDEM-SALYGDYPNVYIEI 160
 QY 156 ANEWFG--SWEQDAMADGY-KQALPRLNAGLNTLAVDAGMGQFQSIHDYREYFN 211
 DB 161 ANEPGHNVWV--DSHIFPAVEEYIPVIRANDPNRIYIVGATWS---QDVHEADNQLD 215
 QY 212 ADPQRTMFSIHMYEYAGNASOVRTINIDRYLNQDLALVIGFGRHTNGD-----VDEAT 267
 DB 216 -DP--NMYAFAHF--YAGTHGOQLRDQYDALSRGAAIFVSEMGTSAAITGDGVFLDEQ 270
 QY 268 I-MSYSEQGVGWLMAS-----WKGNGPEMEYIDLSDNAGNLTANGNTIYNG 315
 DB 271 VWIDFMDERNTSNMWSLTHKDESSAALMPGAP-----TGQWTAELSGAFVVE- 322
 QY 316 PYGLRETSRLSTVFTGGSD--GGTSPTLLYDFBGSMQGWTSGL-----SGGFWA 364
 DB 323 --KIRESAATPPSDPTPPSDPDGPEPDPTPPSD-PGEYRAMPDQNIYNEIYTHNGOLMO 379
 QY 365 VTENS 369
 DB 380 AKWMT 384

RESULT 7
 GUN2_BACS4 STANDARD; PRT; 409 AA.
 AC P06565;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE).

GN CELL.
 OS *Bacillus* sp. (strain N-4).
 CC Bacteria; Firmicutes; *Bacillus*/Clostridium group;
 CC *Bacillus*/staphylococcus group; *Bacillus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87056924.
 RA Fukumori F., Sasahara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RL *Bacillus* sp. strain N-4 and their strong homology."; *Bacteriol.* 168:479-485(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOMYDOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14729; AAA22299.1; -.
 CC PIR: B25156; B25156.
 DR INTERPRO: IPR001547; -.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase.
 FT ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 409 AA; 45690 MW; 9888650B6B4DA3F CRC64;

Query Match 8.7%; Score 226; DB 1; Length 409;
 Best Local Similarity 23.2%; Pred. No. 1.7e-08;
 Matches 104; Conservative 62; Mismatches 147; Indels 136; Gaps 24;

QY 6 KRISITSLILASSILFVSGTSTA-----NANSGFVSGTLLYDANGNPFVWRI-N 57
 DB 2 KKITITFIVLMTLA-LFIIQNTAAADYVVEHGGOLISNGELVNDRGEPVOLKGMSS 60
 QY 58 HGHANYKDATTALBGIANTGANTVRIYLSDGQWTKD-----IHT----- 99
 DB 61 HGLQWYG-----QFVNYESMKMLRDWMTITVFRAAMYTSSGGYIEDP 102
 QY 100 -----VRLISLADNHLVAPEVHDANGYDSIASLNRAVDYIEMRSALIGKEDYIINI 154
 DB 103 VKREYKAEVAEALDGIYIIDWHLISNDPNRIYKEAKEPFDEM-SELGYDYPNVYIE 161
 QY 155 IANEFWGS--WEGD--AMADGYKQALPRLNAGLNTLAVDAGMGQFQSIHDYREY 209
 DB 162 IANPENGSDYVNDQIKRYAE--EVLPIYIRNNDPNRIIY--GTGWSQDVH-HAADN 214
 QY 210 FNADPQRTMFSIHMYEYAGNASOVRTINIDRYLNQDLALVIGFGRHTNGD-----VDE 265
 DB 215 QLTDP--NMYAFAHF--YAGTHGOQLRDQYDALDQGAIFVSEMGTSAAITGDGVFLDE 270
 QY 266 ATI-MSYSEQGVGWLMAS-----WKGNGPEMEYIDLSDNAGNLTANGNTIY 313
 DB 271 AQVWIDFMDERNTSNMWSLTHKDESSAALMPGASP-----TGQWTAELSGAFV- 322
 QY 314 NGPYGLRETSRLSTVFTGG-----SDGQT-----SPTLLYD 345
 DB 323 -----REKIRESAATPPSDPTPPSDPDGPEPDPTPPSDPDYRAMPDNTIYT 376
 QY 346 FE-----GSM--OGWTSGLSG--GFW 363
 DB 377 DEIVYHNGOLWQAKWMTQNDPEPDYGPW 405

RESULT 8

GUNV_ERMCA ID GUNV_ERMCA STANDARD: PRT: 505 AA.

AC 047096: 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE ENOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V) (CELLULOSE V).

GN CELV.

OS Erylnia carotovora.

CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SCRI193;

RX MEDLINE: 94067016.

RA Cooper V.J.C., Salmond G.P.C.;

RT "Molecular analysis of the major cellulase (Celv) of Erylnia carotovora: evidence for an evolutionary 'mix-and-match' of enzyme domains."

RL Mol. Genet. 241:341-350(1993).

CC -1- FUNCTION: ENOGLUCANASE WITH SOME ENOGLUCANASE ACTIVITY. THE PH OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES CELSIUS.

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: X76000; CAA53592.1; -.

DR HSSP: Q06851; INBC.

DR INTERPRO: IPR001547; -.

DR INTERPRO: IPR001956; -.

DR PFAM: PF00942; CBD_3; 1.

DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Cellulose degradation; Hydrolase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 505 ENDOGLUCANASE V.

FT DOMAIN 32 334 CATALYTIC.

FT DOMAIN 333 352 LINKER.

FT DOMAIN 333 505 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

FT SEQUENCE 505 AA; 54900 MW; DBBA9337BB4D2623 CRC64;

Query Match 8.5%; Score 220.5; DB 1; Length 505; Best Local Similarity 21.8%; Pred. No. 5.3e-08; Matches 116; Conservative 69; Mismatches 209; Indels 137; Gaps 21;

QY 2 NNGFKFISITSLASSLTFVSGSTANANG-YYVSGTLYDANGNPFVARGI-NHG 59

DB 6 NQIVRLITLGVYTTVIGMSISFALSATVEYTGQSIENGRVLDQGRVQLGSSHG 65

QY 60 HANYKDOATTAEGIANGTAVRIYLSDGQGTAKD----- 96

DB 66 LQFEGD-----YVKNQSMKVLRODQWGINVRVAMTAADGYISNPUL 107

QY 97 IHTVRNLISLAEDNLHVAPEVDATGYDSIASLNAADVWIMKRALGKETVYINIA 156

DB 108 ANKVKAEVAAQSLGYIITIDHILSDNDPNITKAQAKTFEAFM-AGLGSSPNVITYELA 166

QY 157 NEWFG--SWEGD--AWADGYKQAIPLRLNAGLHNTLWVDAAGGQFPQSIHDYGREVFA 212

DB 167 NEPNGVATWNGQIRPYA---LEVTDTIRSKOPDLITV---GIGTWQDILHD---AADNQ 217

QY 213 DPOHNTFESIMYEYAGNMAQVNTNIDRYLNDLALVIGFGRHTNGD-----VDEAT 267

DB 218 LPDNTMYALHF--YACTHQGLFDRIDYAGSRGALFVSEWGTSDASGPPLPESQT 275

QY 268 IMYSSEQRCVGLWAMSKNGKNGPEWEYL-----DLSNDNAGNLTMGWTYNGPYGLAETS 323

DB 276 WIDFLNNGVSWVMWSLTDKSEASALAPGASKSGWTEONLSTGKFXV-----REQI 328

QY 324 RLSTVFEGGSDGTSPTTLDPEFGSMQGTSSLSGPNVATWMSKSGSHLRADILS 383

DB 329 R-----AGANLGGSDITPTTPEPTNPGNGITG-----DYYL- 359

QY 384 SNSQHYLVHONTSLQONSRIQATVKRANMGVSG--NGMTARLYVTGHTWYSGSFV 440

DB 360 ---QYRVNDNNS---DDAIRMAVNIKNTGSTPIKLSDLQVRYFHD----- 400

QY 441 PINSSGTTSLDLSNQNLSQVREIGVQFQASD-----SSGQTSI 482

DB 401 -DGRPGANLFPVDAWANGPNNTVSTGTPASTDKANRYLVTFSSGAGSL 449

RESULT 9

GUNV_ERMCA ID GUNV_ERMCA STANDARD: PRT: 444 AA.

AC 059394; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE ENOGLUCANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE N) (CELLULOSE N).

GN CELN.

OS Erylnia carotovora.

CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATROSEPTICA FCBR C18;

RX MEDLINE: 98299944.

RA Olesen O., Thomsen K.K., Weber J., Duns J.O., Svendsen I., Wegener C., von Wettstein D.;

RT "Transplanting two unique beta-glucanase catalytic activities into one multienzyme, which forms glucose."

RL Biotechnology 14:71-76(1996).

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: L39788; AAC37033.1; -.

DR INTERPRO: IPR001547; -.

DR INTERPRO: IPR001956; -.

DR PFAM: PF00942; CBD_3; 1.

DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Cellulose degradation; Hydrolase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 444 ENDOGLUCANASE N.

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

FT SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

GN BGIC OR GLD.
 OS Bacillus subtilis.
 CC Bacteria: Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DLG;
 RX MEDLINE: 87194581.
 RA Robson L.M., Chambliss G.H.;
 RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG";
 RL J. Bacteriol. 169:2017-2025(1987).
 CC -1- CATALYTIC ACTIVITY: ENDODIHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M16185; AAA22496.1; ALT_INTT.
 DR PIR: A26874; A26874.
 DR HSSP: Q06851; INBC.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001956; -
 DR PFAM: PF00942; CBD_3; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
 KW SIGNAL
 FT CHAIN 1 29
 FT ACT SITE 169 499
 FT ACT SITE 257 169
 FT DOMAIN 350 499
 FT SEQUENCE 499 AA: 55187 MW: 339D04EE95A65BE1 CRC64;
 SQ
 Query Match 7.2%; Score 188.5; DB 1; Length 499;
 Best local Similarity 21.1%; Pred. No. 8e-06;
 Matches 78; Conservative 75; Mismatches 133; Indels 83; Gaps 17;
 QY 6 KRIFSTLSLLASSILF-----VSGTSTANNSG-FYVSGTLYTDANGNPPVWNG 55
 DB 2 KRISIFITCLIAVLTGMGLPSPASAGTKTPVAKNGOLSIKQTVNRDQKAVOLNG 61
 QY 56 I-NHGAWKQDAITALEGIANTGANTVRIYLSDDGQWTRKD-----IHT----- 99
 DB 62 ISHGLQWGD-----FVKNDSIAKWLDMGIVTFRRAATYTDAGYI 103
 QY 100 -----VNLISLAEDNHLVAPEVHDATGYDSIASINRAVDYIEMRSALIGEDTV 151
 DB 104 DNPVANKVKEAVEAKKEGIYIIDWHILNDGNPNQNKKEKFEKEM-SSLYGNTPVV 162
 QY 152 ITIANEMFG--SWEGD--AMADGYQAIPLRNAGINHTIAMDAAGQOFPSIDYDR 207
 DB 163 IYEIANEPNDVWKKRIKPYAE--EVSIVIRKNDPDNIIV--GTGWSQDVND--- 213
 QY 208 EVFNADPQ--RMTFESIHYEYAGNASQVNTIRIVLNODLALVIGEGHRTNGD-- 262
 DB 214 ---AADDQLKDAVYALH--YAGHGSGLRKNALSKGAPDIVTEGTSDAAGNG 268
 QY 263 --VDEA-TINSYSEQRGVGLANSWKGNGPEWEL---DLSNDWAGNNITANGTIIVG 315
 DB 269 VFLDQSRFNLVLDKSNISVWNWNLSDKQSSALKPKGASKGTGWPPLDITLJAGTFRVN 328
 QY 316 PYGIRERSR 324
 DB 329 IRGTRDSTK 337

RESULT 12
 ID GUN2_BACSU STANDARD; PRT; 499 AA.
 AC P10475;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CARBOXYMETHYL-CELLULOSE) (CMCase) (CELLULOSE).
 GN BGIC OR GLD OR EGLS.
 OS Bacteria: Firmicutes; Bacillus/Clostridium group;
 CC Bacteria: Firmicutes; Bacillus/Staphylococcus group; Bacillus.
 CC [1]
 CC SEQUENCE FROM N.A.
 RC STRAIN-PAP115;
 RX MEDLINE: 87066783.
 RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
 RA Moravell F., Seligy V.;
 RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene";
 RL Nucleic Acids Res. 14:9159-9170(1986).
 CC [2]
 CC SEQUENCE FROM N.A.
 RC STRAIN-CK-2;
 RX MEDLINE: 95225656.
 RA Lindahl V., Aa K., Tronsmo A.;
 RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
 RT subtilis CK-2";
 RL Antoine Van Leeuwenhoek 66:327-332(1994).
 CC [3]
 CC SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Rose M., Entian K.;
 RT "Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases."
 CC [4]
 CC SEQUENCE OF 30-45.
 RC STRAIN-CK-2;
 RX MEDLINE: 95225655.
 RA Aa K., Flensburg R., Lindahl V., Tronsmo A.;
 RT "Characterization of production and enzyme properties of an
 RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
 RT compost soil";
 RL Antoine Van Leeuwenhoek 66:319-326(1994).
 CC -1- CATALYTIC ACTIVITY: ENDODIHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z29076; CAA82317.1; -
 DR EMBL: X04689; CAA28392.1; -
 DR EMBL: X67044; CAA47429.1; -
 DR EMBL: Z73334; CAA97610.1; ALT_INTT.
 DR EMBL: Z99113; CAB13696.1; ALT_INTT.
 DR PIR: A26114; A26114.
 DR HSSP: Q06851; INBC.
 DR SUBTILIST: BG10437; BGIC.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001956; -
 DR PFAM: PF00942; CBD_3; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
 KW SIGNAL 1 29
 FT CHAIN 30 499
 FT ENDOGLUCANASE.

FT ACT SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 FT CONFLICT 283 283 S -> N (IN REF. 2).
 SQ SEQUENCE 499 AA; 55287 MW; 8F735FF71B3BAE2 CRC64;

Query Match 6.9%: Score 181; DB 1; Length 499;
 Best Local Similarity 19.3%; Pred. No. 2.6e-05;
 Matches 100; Conservative 102; Mismatches 188; Indels 128; Gaps 26;

6 KRIFFITSLIASSILF-----VSGTSTANANG-FVSGTTLTDANGPFVANG 55
 2 KRISITITCLITLITLTMGMASAPASAGTTPYAKNQSLIKGTQVNRGKAVQK 61
 56 I-NHGAMTKQAT-TALEGIANGTAV--RIVLSDGGQTKDI-HTVENLISLAE 109
 62 ISSHGLQWGEVYNKDSKMLKDDMGITVFRAAMTTADGQYIDNPSVKKREAVEA 121
 110 NMLVAVPEVHDATGYDSTASLRAYDWTMRMSALIGKEDYIINIANWFG--S 166
 122 LGITVIMHILINDGNPNONKEKAEFKEM--SLYGTNPVITYELANEPNGDV 180
 167 -AMADGYKQALPRLNAGINHTIATDAAGWQFQSIHDYGEVFNADPQ--RTM 222
 181 KPYAE---EVSIVIRKNDPNTIY---GTGWSODVND-----AADDQLADAN 228
 223 HMYETAGNASOVRTNIDRVLNQDLALVIGFGRHRTNGD---VDEA-TINSY 277
 229 HF--YAGTHGQFLRDKANALSKGAPIFTEWGTSDASGNGVFLQSHREMLK 286
 278 GWLAWKNGPMEYEL---DLSNDAGNNLTANGNTIVNPGYGLRESR----- 324
 287 SWVNNWLSDKOESSALKRPGASKTGWRLSDLSASGTFRNILLTKSTKDIPT 346
 325 -----LSTVFTGGSDGTSPTLTYDESGMGWTSGLSGPWAVLTMSKGS 377
 347 KPTQNGISVQYRAG-----DGSN-----NSNQIR 371
 378 ADIQLSN-----SOHYLAVIONTSLQONSRIQATVHANWGVNGMTARLY 426
 372 POLQIKNNNGNTVLDKDYARFYKAKNNG--QNF-----DCDYAQIGKNT 422
 427 --KTHGHTWY-----SGSFYPINGSSGTSLDLSN 456
 423 LHKPGADTYELGFKNGETLAP--GASTGNTQLRLHN 458

RESULT 13
 GUN5_THREU STANDARD; PRT; 466 AA.
 AC Q01786;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4)
 DE (CELLULOSE E-5) (CELLULOSE E5).
 GN CELE.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptococcaceae; Nocardiothermophilaceae; Thermobifida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YX;
 RX MEDLINE; 91258320.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RT "DNA sequences of three beta-1,4-endoglucanase genes from
 Thermomonospora fusca.";
 RL J. Bacteriol. 173:3397-3407(1991).
 RN [2]
 RP REVISIONS.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 DT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

BN [3]
 RE SEQUENCE OF 137-142 AND 157-166.
 RC STRAIN-YX;
 RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
 RT "Activity studies of eight purified cellulases: specificity,
 RT synergism, and binding domain effects."
 RL Biotechnol. Bioeng. 42:1002-1013(1993).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; L01577; AAC09379.1; .
 DR PIR; C42360; C42360.
 DR HSSP; P07986; 1EXH.
 DR INTERPRO; IPR001547; .
 DR INTERPRO; IPR001919; .
 DR PFWAM; PF00553; CBD_2; 1.
 DR PFWAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR KX Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT SIGNAL 1 36
 FT CHAIN 37 466 ENOGLUCANASE E-5.
 FT ACT SITE 299 299 PROTON DONOR (BY SIMILARITY).
 FT ACT SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 466 AA; 49800 MW; 1CF0ADBF2DEF82E CRC64;

Query Match 6.9%: Score 180; DB 1; Length 466;
 Best Local Similarity 25.0%; Pred. No. 2.8e-05;
 Matches 82; Conservative 49; Mismatches 129; Indels 68; Gaps 19;

26 GSTANANSG-FVSGTTLTDANGPFVANGIN-HGAMTKQAT-TALEGIA-WTGAN 81
 160 GTGTVEYKGYKQVCGTQLCDEHGNPVDRKSTIGIOMFDHCLTDSIDLATYWK 219
 82 VRI---VLSDGGQWT-----KDDIHTVRNLISLAEDNMLVAVPEVDATGY 134
 220 IRLSMYIQEDGYETNPRGFTDPMH---QLIDMARTAGLYIVDMILRPGDHY 276
 135 DWIMRSNLIKEDTYIINIANEWFSGWEGDAD--GY-KQALPRLNAGLNHTL 191
 277 TFFAIAORHASKIN-VLEYELANE-----PNGVSASIKSYAEVLPVIRQ 331
 192 AAGNQFQSHIDYREVFNDP--QRMTFSIHMYETAGNASOVRTNIDRVLNQ 249
 332 TRGWSSLGVS-EGSGPAAIAPNPNASINMTAFHY-----AASHRONYL 384
 250 ---VIGFGRHRTNGD-----VDEATIMSYSQGVGLAWSK----- 285
 365 PFVFTERTGTYTIGDANDROMADR--YIDLMARKIGMTKNYSDDFRSAGV 442
 286 -GNGPEWEYLDISNDWAGNNLTANGNTI 312
 443 ASGGP-----WSGSSLSKASGV 460

RESULT 14
 GUN3_BACSU STANDARD; PRT; 499 AA.
 AC P23549;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)

```

DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
GN BGIC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BSE616;
RX MEDLINE; 91299280.
RA Park S.H., Kim H.K., Pack M.Y.;
RT "Characterization and structure of the cellulase gene of Bacillus
RT subtilis BSE616."
RL Agric. Biol. Chem. 55:441-448(1991).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D01057; BAA00859.1; -.
DR PIR; J0111; JN0111.
DR HSSP; 006851; INBC.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR001956; -.
DR PFM; PF00942; CBD_3; 1.
DR PFM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; signal.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55169 MW; 2E821E3D8BBA04 CRC64;
Query Match 6.6%; Score 171; DB 1; Length 499;
Best Local Similarity 19.0%; Pred. No. 0.00013;
Matches 98; Conservative 104; Mismatches 189; Indels 126; Gaps 25;
QY 6 KRIFSTILLASILF-----VSGTANANGS-FYSGTTLXDANGNPFVNRG 55
DB 2 KRISITICILITLTMTMGMLASPASAGCTKTPVAKNGOLSTKQVYNROGKAVQLG 61
QY 56 I-NHGAWTKDQT-TAIGCIANTGANTV--RIVLSDGGQWTKDDI-HTVNLISLAD 109
DB 62 ISSHGLQWGEYVKNLSKLWLRDWDGIVFRAMYTADGIIIDNPVKMKKEAVEAKE 121
QY 110 NHVYAPVPHDANGYSINSLANAVYNTIEMNSALIGKEVTYIINTANEMFG--SNEG 166
DB 122 LGIVYIIDHILNDGNPNONKEKAEFEKM--SLLGNTPNVYIETLANEPNGVNRKDI 180
QY 167 -AMADGKOAIPRLNAGLHNTLMDVDAAGVGPPOSIDHYGRFVNADPO---RNTMFS 222
DB 181 KPAE---EYISIRKNDPNITIV--GTGWSQVND-----AADOLKANAMDL 228
QY 223 HMYEYAGNASOVRTNIDRYLNDLALVIGEFGRHTNG---VDEA-TIMSYSGRGV 277
DB 229 HF--YAGTHGQFLRDRANVALSKGAPLFEVEMGTSASGAGVFILQSRMLKYLDSKI 286
QY 278 GWLAWMKNGKGPWEYL----DISNDWAGNLLAMGNTIYNGYGLREISR----- 324
DB 287 SWNNMNLSPKQESSALKPKGASKTGRLSDLSASGTFYRENTIIGKTDIPDPAPND 346
QY 325 -----LSTVFTGGSGDGTSPITLYIDFGSMQGWTSLSLGGPNAVTEWSSKSHSLK 377

```

```

DB 347 KPTGNGISVYRG-----DSM-----NSMDIR 371
QY 378 ADIQLSNSQHYLVHNTLSQ-----QNSRIQATVKHANWGNGMARTLYV-- 426
DB 372 PQLQIKNG-----NTYDLADVTARYWYKMKNGVNDQYALGCGNVTYKRVTL 423
QY 427 -KTGHGTYW-----SGSFYPINGSGSTILSLDLSN 456
DB 424 HKPQAGADTYELGFKNGTLP--GASTGNTQLRLHN 458
RESULT 15
GUND_CLOCE STANDARD; PRI; 584 AA.
AC P25472;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE ENDOGLUCANASE D. PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE D)
DE (CELLULOSE D) (CECCD).
GN CELCDD.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35319;
RX MEDLINE; 92009193.
RA Shima S., Igarashi Y., Kodama T.;
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
RT celCDD, of Clostridium cellulolyticum."
RL Gene 104:33-38(1991).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIOSYLASES THAT CUT THE DISACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- PATHWAY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90341; BAA14354.1; -.
DR PIR; J01229; J01229.
DR HSSP; P17901; IEDG.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002048; -.
DR INTERPRO; IPR002105; -.
DR PFM; PF00404; Dockerin_1; 2.
DR PFM; PF00150; cellulase; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; signal.
FT SIGNAL 1 24
FT CHAIN 25 584 ENDOGLUCANASE D.
FT DOMAIN 25 328 CATALYTIC (BY SIMILARITY).
FT DOMAIN 329 353 PRO/THR-RICH (LINER).

```

FT DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 530 584 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 530 552 1.
FT REPEAT 562 584 2.
FT ACT_SITE 159 159 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 264 264 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 584 AA; 66061 MW; 0FC41257E81322C3 CRC64;

Query Match 6.5%; Score 170.5; DB 1; Length 584;
Best Local Similarity 20.5%; Pred. No. 0.00017;
Matches 119; Conservative 91; Mismatches 204; Indels 167; Gaps 30;

QY 6 KRIFITITLLASLILFVSGSTANANGFVSGTLYDA-NGNPFMR---GINHGH-60
DB 2 KRIILALITSCSTIMSLFPM-----VYGAINSQDVMKKGIGMILNT 44
QY 61 -----AMYKQATTALEGIANGTANTVRIYLSDGQWKDITVRNLTSLAEDNHY 113
DB 45 FDAPFEGSMKAAQRYTDDFKQAGFKHRIPI---RM---DOHTLANSPTVDSNFLN 97
QY 114 AVPEY-----HDATGYDSIA-SLNRAVDYIEMRSALIGKEDVIINIA 156
DB 98 RIEFYIDMSLSRGFTYINSHDFTWLMQNYSONIGREFKIMEQIAQRFKGSENLVFEIL 157
QY 157 NEMFGSMGEGDAMADQYKRAIPRLRNAGLHTLMVDAAGQFPOSIDHYGREVENADPQR 216
DB 158 NEPRGNITDSQLDMDKRLILNIRKTNPTRNVIIGAGYNSY-NSLSQL-EIPN-DP-- 211
QY 217 NTMFESIHYE-YAGNAGSOVRTNIDRVLNODLAVIGFGRHRTNGVDVDEATINSYSOR 275
DB 212 NLIAFFHYIDPYSFTHQ-----GTWG---TKNDMD-AIAMVFNHVK 250
QY 276 GVGWLAMSMKNGPEM-----EYLDLSNDMA--GNNLTAMGNTIVNGP 316
DB 251 -----KMSDKNNIPYILGEYGVMGSHDRSAVKMFEDFVSDQALSHGFCGAMD---NGV 301
QY 317 YGL-----RETSR-----LSVFTGGSGDGTSP-----TLX 344
DB 302 FGSVDNDMAFYNRDRTPQDKETILNAILTGTGYDTPPTETNPDPPTAPATPAYGEOLIE 361
QY 345 DEEGSMQGTGSSLSGGFWAVTEWSSKSGSHKADIDQLSSNSOHTLAVIYONTSLQON--- 401
DB 362 DFEGLMQ-W--AAYSGVATASCKISSKSNNGLEITYAGSSNGYGVNDENHRQDMEX 418
QY 402 -SRIQATVYKHAMWGSVNGMTARLVKTGHGYTWYSGSFVPINGSSGTTLSLDSN--- 456
DB 419 WOKISFDIKSNTNEVRLLIAEQSKRIEGEDGBHW---TYVILKPSWTITIEIPFSSFTKR 475
QY 457 -----VQNLQVREI-----GVQFOSASDSSGQTSIYIDNV 487
DB 476 MDYQPPADGSETFDLYKVGSLHFYATSNNSG--TLNIDNI 514

Search completed: December 19, 2000, 16:44:15
Job time: 891 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:00:19 ; Search time 363.46 Seconds
(without alignments)
175.883 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNGRKRTISITLSLMASS.....QSASDSGQSTIIDVIVE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_14:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1622	62.2	516	2 066185	066185 bacillus cl
2	814.5	31.2	395	2 069347	069347 vibrio sp.
3	761.5	29.2	384	2 086399	086399 streptococ
4	634	24.3	279	2 092F13	092F13 thermomon
5	628	24.1	930	2 09RFK5	09RFK5 calidbacill
6	541	20.8	327	2 09RJ15	09RJ15 streptococ
7	253.5	9.7	1097	2 09ZAI7	09ZAI7 anaerobic t
8	239.5	9.2	557	2 P94622	P94622 clostridium
9	228	8.7	389	2 059232	059232 bacillus sp
10	203.5	7.8	570	2 059665	059665 pseudomon
11	193	7.4	499	2 045532	045532 bacillus su
12	187.5	7.2	476	5 018453	018453 heterocera
13	187.5	7.2	481	2 083012	083012 actinomyces
14	187.5	7.2	501	2 031029	031029 erwinia sp
15	186.5	7.2	387	2 031029	031029 erwinia car
16	185.5	7.1	930	2 059290	059290 clostridium
17	180	6.9	486	2 045430	045430 bacillus sp
18	180	6.9	635	2 066065	066065 fibrobacter
19	179	6.9	783	2 045554	045554 bacillus sp

20	178	6.8	749	2 059154	059154 anaerocellu
21	177	6.8	499	2 052731	052731 bacillus sp
22	173	6.6	621	2 007653	007653 cellvibrio
23	169	6.5	821	2 059241	059241 bacillus sp
24	161	6.2	478	5 016028	016028 globodera r
25	160	6.1	278	2 09RK56	09RK56 streptococ
26	152	5.8	506	5 09UA57	09UA57 meloidogyne
27	148.5	5.7	470	5 09UM65	09UM65 globodera t
28	145.5	5.6	395	5 09UM64	09UM64 globodera t
29	140.5	5.4	1335	12 089353	089353 parametium
30	138.5	5.3	910	3 P87211	P87211 orpinomyces
31	136.5	5.2	494	2 086039	086039 alteromonas
32	136	5.2	438	3 09URL8	09URL8 candida alb
33	133.5	5.1	319	5 018454	018454 heterodera
34	132	5.1	1300	2 032591	032591 escherichia
35	132	5.1	1300	2 032555	032555 escherichia
36	131.5	5.0	319	5 077449	077449 heterodera
37	130.5	5.0	319	5 061595	061595 heterodera
38	130	5.0	392	5 044078	044078 globodera r
39	129	4.9	391	5 077094	077094 globodera r
40	129	4.9	519	2 047916	047916 fibrobacter
41	129	4.9	1070	2 055365	055365 synechocyst
42	129	4.9	1107	2 P76088	P76088 escherichia
43	129	4.9	1167	2 P76859	P76859 escherichia
44	129	4.9	1536	2 048031	048031 haemophilus
45	127.5	4.9	739	2 09X687	09X687 salmonella

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	516 AA.
066185	AC	066185;		
066185	DT	01-AUG-1998 (TREMBLrel. 07, Created)		
066185	DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
066185	DT	01-MAR-2000 (TREMBLrel. 13, Last annotation update)		
066185	DE	MANNANASE.		
066185	OS	Bacillus circulans.		
066185	OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
066185	CC	Bacillus/staphylococcus group; Bacillus.		
066185	RN	[1]		
066185	RP	SEQUENCE FROM N.A.		
066185	RK	MEDLINE; 9823274.		
066185	RA	Yoshida S., Sako Y., Uchida A.;		
066185	RT	"Cloning, sequence analysis, and expression in Escherichia coli of a		
066185	RT	gene coding for an enzyme from Bacillus circulans K-1 that degrades		
066185	RT	guar gum."		
066185	RL	Biosci. Biotechnol. Biochem. 62:514-520(1998).		
066185	DR	EMBL; AB007123; BAA23878.1; -.		
066185	DR	INTERPRO; IPR001547; -.		
066185	DR	PFAM; PF00150; cellulase; 1.		
066185	DR	SEQUENCE 516 AA; 55245 MW; DBE5A48842AE8062 CRC64;		

Query Match	62.28;	Score 1622;	DB 2;	Length 516;
Best local similarity	59.68;	Pred. No. 1.2e-107;		
Matches 292;	Conservative 86;	Mismatches 84;	Indels 28;	Gaps 2;
QY	27	TSTNANSGFVSGTLYDANGNPFVNGINHGAWYDQATTEIGTANGANTVRVL	86	
DB	29	TNRKHAASGFVSGTKLDAATGQPPVKGVNHATWYDQSTALPAKTAGANTIRVL	88	
QY	87	SDGQWTKDDIHTVNTNLSLAEDNHLVAVPEVHDATGDSIASINRAVDYIEMRSALIG	146	
DB	89	ANGHWKTDVNTVNTNNTLLOQNKLIIVLEVHATGSDSLSDLDNANNVYIGIKSALIG	148	
QY	147	KEDVYIINAEWFGSGWEGDAMADGYKQAIPLRNAGINHTLWADAGGCFPOSIDHYG	206	
DB	149	KEDRYIINAEWYGTWGVAMANGYKQAIPLRNAGITHTLWADAGGCFPOSIVKNYG	208	
QY	207	REVNADQQRNTWFSIHNYEYAGNAGSVRTNIDRYVNLQDLALVIGEGHHTNGDVDEA	266	

```

DB 209 TEVLNADPLAKNTVFSTHMEYAGNASTYKSIDVGLKNNLLIGEGCGHTNGVDRA 268
QY 267 TMTSSEORGVGWLAMSKNGPEWEYIDLSDNMGANNLTANGNTVNGPYGLRETSRLS 326
DB 269 TMTSSEORGVGWLAMSKNGPEWEYIDLSDNMGANNLTANGNTVNGPYGLRETSRLS 328
QY 327 TMTSSEORGVGWLAMSKNGPEWEYIDLSDNMGANNLTANGNTVNGPYGLRETSRLS 360
DB 329 GTF--GGTPTTSSPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPT 386
QY 361 GFWATVEMSSKSGSHKADIDQLSSNOHYIQTSTLQONRIQATVYKHAMSGVNGM 420
DB 367 GFWATVEMKATGAQTLKADVLSQNSHSLYITSNQNSKSLKATYKHAMSGVNGM 446
QY 421 TARKLVKGGHTVSGSVVPFNGSSGCTTSLDLSNVONTLSOVRIGVQFOSGSSGCT 480
DB 447 YAKLVYKGGHTVSGSVVPFNGSSGCTTSLDLSNVONTLSOVRIGVQFOSGSSGCT 506
QY 481 SIYDNVIVE 490
DB 507 AIYDVSLSQ 516

```

```

RESULT 2
ID 069347 PRELIMINARY; PRT: 395 AA.
AC 069347;
DT 01-AUG-1998 (TREMREL. 07, Created)
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
DE 01-MAY-2000 (TREMREL. 13, Last annotation update)
GN BETA-1,4-MANNANASE.
OS Vbrio sp.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MA-138;
RA Tanaru Y., Ataki T., Morishita T., Kimura T., Sakka K., Ohmura K.;
RL J. Ferment. Biolog. 83:201-205(1997).
DR EMBL: D86329; BAA25188.1; -
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR002883; -
DR PFAM: PF00150; cellulase; 1.
DR PRAM: PR02013; CBD_5; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 395 AA; 43098 MW; A3E30B3A978C5DA1 CRC64;

```

Query Match 31.28; Score 814.5; DB 2; Length 395;
 Best Local Similarity 46.58; Pred. No. 2.3e-50;
 Matches 168; Conservative 62; Mismatches 104; Indels 27; Gaps 9;

```

QY 6 KKSITLSTLLASSILFVSGSTANANSGFYVSGTITLYDANGNPFVKGNGINHGHWYKD 65
DB 5 KALSTL-----LFTWASC---AHAGFYVSGVLYTEANGSAFKRGINHAHTWTD 52
QY 66 QATTAIEGANTGANTVRYVLSGCGWTKDDITVNRNLISLAEDNHLVAVEVDATGY- 124
DB 53 KLSVALSGIATGANTVRYVLSNGYRWTKNDVDTNINAKANNLTAILEVDITGYG 112
QY 125 --DSIASLNRAVDYWIEMKSLIGKEDVTIINANWFG--SWEGDANADGYKQAIPLRN 181
DB 113 EESSASISDSADYWIEMKSLIGKEDVTIINANWFG--SWEGDANADGYKQAIPLRN 172
QY 182 AGLNHTLVNDAAGWQ--FPOSIDYGREVENADPQRTMSEIHME--YAGNASQVRTNI 239
DB 173 AGLNHTLVNDAAGWQ--FPOSIDYGREVENADPQRTMSEIHME--YAGNASQVRTNI 230
QY 240 DRYLNODLALVYIGFGRHTNGVDVDEATIMSYSEORGVGLAMSKNGPEWEYIDLSDN 299
DB 231 SSFTNNGLVLYIGFGRHTNGVDVDEATIMSYSEORGVGLAMSKNGPEWEYIDLSDN 290

```

```

QY 300 WAGNNLWAGTIVNGPYGLRETSRLSTVFTGG---SDGCTSP---TTLYDEFGSMOG 352
DB 291 WDNNSYSTWGVVLLINGONGIKSTLTATVFCGNDCCDDSSGGEYPCSSAYVDDGCGW 350
QY 353 W 353
DB 351 W 351

```

```

RESULT 3
ID 086599 PRELIMINARY; PRT: 384 AA.
AC 086599;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DE 01-MAY-2000 (TREMREL. 13, Last annotation update)
DE SECRETED BETA-MANNOSIDASE.
GN MANA2.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinash H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL031514; CA20610.1; -
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR002883; -
DR PFAM: PF00150; cellulase; 1.
DR PRAM: PR02013; CBD_5; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 384 AA; 39847 MW; B8AD158E3FD3BD11 CRC64;

```

Query Match 29.28; Score 761.5; DB 2; Length 384;
 Best Local Similarity 44.28; Pred. No. 1.3e-46;
 Matches 161; Conservative 59; Mismatches 121; Indels 23; Gaps 8;

```

QY 9 FSITLSTLLASSILFVSGSTANANSGFYVSGTITLYDANGNPFVKGNGINHGHWYKD 65
DB 16 FAAYLGLFLA-----LAGSAGRAEAAAGSIHNSNGVLENGSVYMRGVNATYTPD 70
QY 66 QATTAIEGANTGANTVRYVLSGCGWTKDDITVNRNLISLAEDNHLVAVEVDATGY- 124
DB 71 R-TGSIADIAKANTVRYVLSGCGWTKDSASEVSALIGQCANRYVICLVEHDTGYG 129
QY 125 --DSIASLNRAVDYWIEMKSLIGKEDVTIINANWFG--SWEGDANADGYKQAIPLRN 182
DB 130 EDGAATSLDAAAYWYSKALSGQEDVYVYVNGNEPFTNTTAWTDTAKSALIGLUGA 189
QY 183 GLNHTLVNDAAGWQ--FPOSIDYGREVENADPQRTMSEIHME--YAGNASQVRTNI 241
DB 190 GLDHALVNDAPNNGQWMSGRMSNAASVFASDPDRMTVSVMHYG--YDTPAAVRYLNA 248
QY 242 VLNODLALVYIGFGRHTNGVDVDEATIMSYSEORGVGLAMSKNGPEWEYIDLSDN 301
DB 249 FVSGSLPIVYIGFGRHTNGVDVDEATIMSYSEORGVGLAMSKNGPEWEYIDLSDN 308

```


09RJ15 PRELIMINARY; PRT; 327 AA.
 AC 09RJ15: TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)
 DE PUTATIVE SECRETED BETA-MANNOSIDASE (FRAGMENT).
 GN MANA.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomyces; Streptomyetaceae; Streptomyces.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parthill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinsahl H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL121746; CAB57406.1; -.
 DR INTERPRO: IPR001547; -.
 DR PFM: PFO0150; cellulase; 1.
 FT NON_TER 327
 SQ SEQUENCE 327 AA; 34772 MW; A21B068C2C0975D CRC64;

Query Match 20.8%; Score 541; DB 2; Length 327;
 Best Local Similarity 41.5%; Pred. No. 5e-31;
 Matches 117; Conservative 54; Mismatches 101; Indels 10; Gaps 8;

QY 25 SGGSTANAN-SGFFVSGTLLYDANGNPFVARGINGHAWTKDOATTAIEGANTGATVR 83
 DB 47 AGPSSAGAGAGAGHIGDGRLEBNGNDFVARGVNHATWPGF-TQSLADYKALGNNSVA 105
 QY 84 IVLSDGQWTKDIDHYVRLISLAEDNHLVAPEVDATGY--DSTA-SINRAVDYIEM 140
 DB 106 VVLSDDGHRGSENGPADVAAYIEQCKANRLICVLEHDTGTGAEDAAAGTIDHADYWIGL 165
 QY 141 RSLALIGEDVIYININEMGSMGDMADGYQALPRLNAGINHTLMDAAGMGQFPQ 200
 DB 166 KDVLAGEDEVIVINIGEPNGNDPAGWTEPYAAVAKLLAAGLQHTIMVDAFPMGQDWQ 225
 QY 201 SI-HDYGREVENADPORHTFESIMHYEAGNNAQVNTNIDRYLMDLALVIGEF-GHRH 258
 DB 226 GVMRANRASYDADPTGNLIFSITHYS-VPTAQEITDYINAFVADALPILIGFEGPAD 284
 QY 259 TNGDVDEATMTSSEORGCVGLANSWKNGPEWETLDSMDW 300
 DB 285 QYDPPDEDTMTATAEQRLGLYLAWSMGND--FVLDLALDF 324
 RESULT 7
 ID 09ZAI7 PRELIMINARY; PRT; 1097 AA.
 AC 09ZAI7: TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, last annotation update)
 DE S-LAYER ASSOCIATED MULTIDOMAIN ENDOGLUCANASE.
 GN CELA.
 OS anaerobic thermophile KM-THCT.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-KM-THCT;
 RX MEDLINE; 99175466.
 RA Cann I.K., Kocherginskaya S., King M.R., White B.A., Mackie R.I.;
 RT "Molecular cloning, sequencing, and expression of a novel multidomain
 RT mannanase gene from the anaerobic bacterium polysaccharolyticum";
 RL J. Bacteriol. 181:1643-1651(1999).
 DR EMBL; U82255; AAD09354.1; -.
 DR INTERPRO: IPR001119; -.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR002088; -.
 DR PFM: PFO0150; cellulase; 1.
 DR PFM: PFO0395; SLH; 3.
 DR PFM: PFO2018; CBD; 6; 2.
 DR PROSITE: PS00904; PPTA; UNKNOWN_1.
 DR PROSITE: PS01072; SLH_DOMAIN; 2.
 SQ SEQUENCE 1097 AA; 119762 MW; C41CEB92C0494B9 CRC64;

Query Match 9.7%; Score 253.5; DB 2; Length 1097;
 Best Local Similarity 22.3%; Pred. No. 7.8e-10;
 Matches 104; Conservative 76; Mismatches 181; Indels 105; Gaps 20;

QY 8 IFSITSLIASILFVSGTSTANNSG--FYVSGTLLYDANGNPFVARGIN-HGH-AW 62
 DB 10 VVAVLVTAALNGVYIGDKQAAAGTSCDGRHYVGNKIYDPDGNDFYIKVGNIGYRSW 69
 QY 63 YDQATTAIEGANTGA-NVRIYVSDG-----GQWTKDHYVRLISLAEDNHLVAV 115
 DB 70 ERSVYLDVHLIADYWKFTVRLNCFIGNMNGEGTGANNIDAI--IKAFYAKVYVE 126
 QY 116 PEVHATGYDSTA-----SLNRAVDYIEMSGALIGKEDYITINANMFSGSWE-- 164
 DB 127 IDLHPTGTGPPISNPPAPGQPSLDQAI-AWTKELAAKYKDNPIYFWNTANEPGSLAPL 185
 QY 165 GDWADGYKQALPRLNAGINHTLMDAAGMGQFQSIHD-----YGREVF 210
 DB 186 DQWYANEEIKAKRSIGADNIIVYD--GWSYANEGIEQNPTVDEKSAVLTGQDGL 243
 QY 211 NADPQNTWFSIMHYEAGNNAQVNTNIDRYLMDLALVIGEFHRTNGVDATIMS 270
 DB 244 NADSAKNTFFAHYNN-EDGIQKVEDYIDRANAGLYFMEYGRKDS--DAAKEGVKS 300
 QY 271 YSE---QRCVGLANSWKNGPEWETLDSNDMAGNNLWAGNTLVNGPYGLRSTSLST 327
 DB 301 GLQAVNNKAGRIYNNWDS---YDYLDTSG----- 328
 QY 328 VFTGGG-----DGTSTPTLYDEGSMQGTGSLSGPVAWTEWSSKGSLSKADIQ 381
 DB 329 --TGRRSGWEINKTGSKPTNL-----SWVGDKIMDNDNGIT--PFDONPKVDLA 376
 QY 382 LSSNSOYLVHVIQNTSLQONSRIQATVKRANNGSNGNTALAYK 427
 DB 377 IER-----LIANNNGFKAGDRVOFTFLRNSGDLPIGDSKRYVK 416

RESULT 8
 ID P94622 PRELIMINARY; PRT; 557 AA.
 AC P94622: TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, last sequence update)
 DT 01-MAY-1997 (TREMBlrel. 03, last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, last annotation update)
 DE ENDO-1,4-BETA GLUCANASE ENGF (EC 3.2.1.4) (CELLULOSE) (ENDOGALUCANASE)
 DE (ENDO-1,4-BETA-GLUCANASE) (CANBOXYMETHYL CELLULOSE).
 GN ENGF.
 OS Clostridium cellulovorans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97136706.
 RX Shewetta S.A., Ichl-Ishl A., Park J.S., Liu C., Malburg L.M.,

RA Doi R.H.:
 RT "Characterization of enf, a gene for a non-cellulosomal Clostridium
 RL cellulovorus endoglucanase";
 CC Gene 162:163-167(1996).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: U37056; AAB40891.1; .
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; .
 DR PFMW: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Hydrolase; Glycosidase.
 KM SEQUENCE 557 AA; 60131 MW; D186EC88E50AED CRC64;
 SQ

Query Match 9.2%; Score 239.5; DB 2; Length 557;
 Best Local Similarity 23.1%; Pred. No. 3e-09;
 Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

QY 2 NNGFKRIFSI-----TSLSLASSILFVSGTS-----TANAN-----SGFTV 38
 DB 3 NNVKRRILSIYAGAMLMALVPINVAETYSNLTGNANVKKPSVGRKQLQINKNGI-- 60
 QY 39 SCTTLYDANGNPFVNRGIN-HGHAWTKDQATAIEGIANTGA-----NTRYIV- 86
 DB 61 --KTLCDKDNFPIQLRGMSTHGLQMF-----PEYVNNNAFALSDNWSNVITLAMY 110
 QY 87 -SDGGQWTKDDI-HVYRNLSIAEDNHLVAVPEVDATGYDSIASINRAVDYWIERSAL 144
 DB 111 VAEQVATNPSYKQVINGINVAIANDMVIVDMHMMNGDNASYSSAQGFENDISLT 170
 QY 145 ICKEPTVITININWFGSGEG--DA--WADGYKQALP--RLRNAGLHETLMDAAGMG 196
 DB 171 YNNNNIITYELCNEPENGSGVTNDATGWAQYSATPIVQLLRKNGENLITVGNPEWS 230
 QY 197 QEPQSIHDGREFVNAADPQRTMFSIHMYEYAGN-ASQVETNIDR-----VLNQDA 248
 DB 231 QRPDLAAD-----NFINDSNTMTSYHF--YSGTNPISTVDNRDNAMSNVYALNHGA 282
 QY 249 LVIGEGFHRTNGDVEDATIMSYSEBQGVWLAMSKNGSPMEYLDLSND----- 299
 DB 283 VFATEMGTSLATG-----TTGPYL-AKADAWLDF-LNGNNISMCNFSINRDEKAALNS 335
 QY 300 -----WAGNNLTWAG-----NTIVGPGGLRRTSKLSTVFTGG 332
 DB 336 LTSIDPGSDKRLMADNELTTSQGYVARIKGAYYAPVPDVTQPTAPKDFSSGFDFNDG 395
 QY 333 GSDG-----GTSPTLYDFEGSGMGWGTGSLSGPNAVITWMSKSGSHLK-----ADIOS 383
 DB 396 TTQGGVNPDSPIYALINVENANNALKISML-----NSKGSNDLSEGNFWANVRIS 445
 QY 384 SNS-OHYLVAVIONTSLOQNSRIQATVYKRAHMGSVNGMTARLYVYKGGHY-----TW 434
 DB 446 ADIMGQINIVYDPTKLTMDVIAPTPV-----NYSIALPQSTHGWGPTAIRVW 496
 QY 435 YSGSVPIINGSGTSLSDLSNVQLSQVREIGVDFQASDS-----SGQTSITI 484
 DB 497 TNNFYAQTDGYKATLTISTNDSPNFTIA-----TDAADSVYTNMILFVSSNDISLT 550
 QY 485 DNV 487
 DB 551 DNI 553
 RESULT 9
 ID 059232 PRELIMINARY; PRT; 389 AA.
 AC 059232
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE ENDO-BETA-1,4-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULASE)
 DE (ENDOGUCANASE) (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE).

OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 RN Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-186-1;
 RA Sanchez Torres J.;
 RL Thesis (1994); Universidad de Salamanca, Spain.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: 233876; CA83942.1; .
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; .
 DR PFMW: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Signal; Hydrolase; Glycosidase.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 389 ENDO-1,4-BETA-GLUCANASE.
 SQ SEQUENCE 389 AA; 43688 MW; 9111397485609B5F CRC64;

Query Match 8.7%; Score 228; DB 2; Length 389;
 Best Local Similarity 23.6%; Pred. No. 1.2e-08;
 Matches 100; Conservative 65; Mismatches 154; Indels 104; Gaps 21;

QY 6 KRFISITSLASSILFVSGTSTANANS-----GFYVSGTTLYDANGNPFVNRGI-NH 58
 DB 2 KRITTFVLLMTVALFSIGNTTAADNSVVEHGLQSTISNGELVNERGEQVQLGMSH 61
 QY 59 GHAWTKDQATNIEGANTVRLYLSDDGQWTKDD-----IHT----- 99
 DB 62 GLQWVG-----QFVNTESKMLRDLMDGIVFRAAMTSSGGYIDPS 103
 QY 100 -----YRNLSIAEDNHLVAVPEVDATGYDSIASINRAVDYWIERSALIGKEDYIINI 155
 DB 104 VAEKTKREAVEAALDDIYIIDMHLSDDNPRIYKEAKPFDEM-SELYGYPVITYEI 162
 QY 156 ANEWFGSWEGDAMADGYK---QALPRLNAGLHETLMDAAGWQGFQPSIHIDYREFVN 211
 DB 163 ANEPNPS--DYTWDRIRKPYAEVIVPIRNDPNIIIV---GTGWSQDVH-HAADNL 216
 QY 212 ADPQRTMFSIHMYEYAGNNAQVRYTNDVRLNODLALVIGEGFHRTNGD-----YDEA 266
 DB 217 ADP--NVMTAFHF--YAGTREGONLDQVDYALDQGAALFVSWGVSATGDSGVFLDEAQ 272
 QY 267 TMSYSEBQGVWLAMS-----WKGNGPMEYLDLSNDVAGNNLTWAGNTIYNG 315
 DB 273 EMIDMDERNLSWAMSLHKDESSAALMPGANP-----TGWTAEALSFGFTVRE- 324
 QY 316 PYGLRSTSLSTVFTGGSDGT---SPTLYDFE-----GSM---QGWGSSLSG--- 360
 DB 325 --KIRSEASIPSPDPTPSPDPEYPAWPTQYLTNEIYVHNGQLWQAKWTONQEPGDPY 382
 QY 361 GPW 363
 DB 383 GPW 385
 RESULT 10
 ID 059665 PRELIMINARY; PRT; 570 AA.
 AC 059665
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE) (ENDOGUCANASE)
 DE (CARBOXYMETHYL CELLULASE).
 GN CELE.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 CC Pseudomonas.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE: 95366948.
 RA Hall J., Black G.W., Ferreira L.M.A., Millward-Sadler S.J.,
 RA Ali B.R.S., Hazlewood G.P., Gilbert H.J.,
 RT "The non-catalytic cellulose-binding domain of a novel cellulase from
 RT Pseudomonas fluorescens subsp. cellulosa is important for the
 RT efficient hydrolysis of Avicel."
 RL Biochem. J. 309:749-756(1995).
 CC -1- CATALYTIC ACTIVITY: ENOCHDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: X86798; CAA60493.1; -
 DR HSP; P07103; 1E62.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001979; -
 DR INTERPRO: IPR002883; -
 DR PFAM: PF00150; cellulase: 1.
 DR PFAM: PF00553; CBD_2; 1.
 DR PFAM: PF02013; CBD_5; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Hydrolyase; Glycosidase.
 SQ SEQUENCE 570 AA; 59925 MW; D00B26F13D57FBE7 CRC64;

Query Match 7.8%; Score 203.5; DB 2; Length 570;
 Best Local Similarity 23.3%; Pred. No. 1.1e-06;
 Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

QY 14 SLLASSILEFSGSTANANGSGFYVGTLLYDANGNPFVARGIN-----HGHWYKD 65
 DB 22 AIIILSFGILGVSNAQADVAPISVGNKIL-ANGOPASGSKLSFMSNTEMGEXYNA 80
 QY 66 QATTAIEGANTGANTVRIY---SPGGQWT---KDDITVYANLISLAEDNHLVAVPY 118
 DB 81 QVVSFWLK---SDMNAKLYRAMAGVEDEGGYLPDANDRYQVVD-AALANDMTIIDMS 137
 QY 119 HDATGDSIASLRAYDWTIEMRSALIGKEDTYINIANEMFG-SWEGDAMADGYKCA-I 176
 DB 138 HNHQJQS---QALFEQEM-ARKYGANNHYIETIENPLQVMSNT--IKPYQAVI 189
 QY 177 PRLNAGLNTLWVADAGMGOPOSTIDYGREVNADP---QRTWFSIMTYEAGSNMS 233
 DB 190 AAIKRAIDPMLIYGTPTWQ-----DVDVANADPIGYONIAATLHF--YAGHGQ 239
 QY 234 QVETNIDRYANQDLALYIGFEGHRTGDDVDENTISYSGRGVGLAMWKGPEWEX 293
 DB 240 YLRKQACTALNRGIALFTVTEGVSANNGD-----GAVANSE-----GW 315
 QY 294 LIDLSNDVAGNNLITAMGWTIVNGPYGLRETSLSTVFTGGSGDGGTPTLLYDEGSMQW 353
 DB 276 ---TNANVSFKYT---KHISNANALNDKYEGASALVPGASANG-----GW 315
 QY 354 TGSSL-SGGPWA---VTEW---SSKGSLSKADLIQSSNQHYLHYQNTSLQONSRIQ 405
 DB 316 VNSQLTASGALAKSLIISGWPSTYNTSSSSSAVSQTOVSSSSQ--APVSSSSSTASSVVS 373
 QY 406 ATY--KHAMGSGVNGMTARLYVKTGHGYT-----YSGSFVPIGSSGTTLSL 452
 DB 374 SAVSGGQCNW---YGLIYPLCSTTTNGMGNENASCIARATCGGAPAPMGIVGSGTSS- 428
 QY 453 DLSNVQNLQVREIGVQFQASDSSGQTSI 482
 DB 429 -----QASSSVRSSSSSLVSSRSSSSSV 453

RESULT 11
 Q45532 PRELIMINARY; PRT; 499 AA.
 AC Q45532;
 DT 01-JAN-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE CELLULASE.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87190397.
 RA Nakamura A., Uozumi T., Teruhiko B.,
 RT "Nucleotide sequence of the cellulase gene of Bacillus subtilis."
 RL Eur. J. Biochem. 164:317-320(1987).
 DR EMBL: M28332; AA422307.1; -
 DR HSP; O85465; 2A3H.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001956; -
 DR PFAM: PF00150; cellulase: 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM: PD001947; -; 1.
 SQ SEQUENCE 499 AA; 55075 MW; D42AAB9E63B94D23 CRC64;

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 5.2e-06;
 Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

QY 6 KRISITLSLASSILEF-----VSGSTANANGS-FYVSGTLLYDANGNPFVARG 55
 DB 2 KRISIFITCLITLITVTFMGILQASPASAGKTPRAKNGQLSINGTOLVNDGKAVOLKG 61
 QY 56 I-NHGHWYKQATTAIEGANTGANTVRIYSPGGQWTKD-----IHT----- 99
 DB 62 ISSHLOWYGD-----FYNKSLKLRDNGKITVFRAMATYADGGYI 103
 QY 100 -----VRNLISLAEDNHLVAVPEVDATGYDSIASLRAYDWTIEMRSALIGKEDTV 151
 DB 104 DNPSTKKNYKAEVAKELGIYIIDHILLNDGNPNQKERAKDFEEM--SLGNGPNV 162
 QY 152 IINIANEMFG--SWEGD--ANADGYKQAIPLRRAGLNTLWVADAGMGOPOSTIDYGR 207
 DB 163 IYEIANEPNGVNMKRDIKPYAE--EVIYVRKNDPNITIV--GTWSDQVND--- 213
 QY 208 EYFNADPO---RNTWFSIMTYEAGSNASQVETNIDRYANQDLALYIGFEGHRTNGD-- 262
 DB 214 ---AADOLKANVYKALHF--YAGTHQGSIRKANVYLSGAPLFYEWGSDASGNGG 268
 QY 263 ---VDEATISYSGRGVGLAMWKGPEW--EYIDLSN--DWAGNNLITAMGWTIVNGPYG 318
 DB 269 VFLDQSR-----EWLNTYDSKNISVWNNL----- 293
 QY 319 LRETSLSTVFTGGSGDGGTPTLLYDEGSMQWTSLSGSPVATWMSKGSLSKA 378
 DB 294 -----SDKOESSALK-----PGASKTGG-WPLIDLTASGTF-VRE 327
 QY 379 DI--QLSSNQHYLHYQNTSLQONSRIQATVYKHAMGSGVNGMTARLYVKTGHGYT--- 433
 DB 328 NILGNNDSTKKEPFPADQNPQENG--ISQYKADGQVNSQIRPOLHIKKNGNATYDL 386
 QY 434 -----WTSGSFVPIGSSGTTLSIDLSNV 457
 DB 387 KDVTARYWYNNAK-----NKGQNFDDYAOI 411

RESULT 12
 O18453
 ID O18453 PRELIMINARY; PRT; 476 AA.
 AC O18453;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE BETA-1,4-ENDOGLUCANASE-1 (CURSOR (EC 3.2.1.4) (CELLULASE)
 DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE).
 GN HG-ENG-1.
 OS Heterodera glycines.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 98226740.
 RA Smant G., Stokkermans J.P.W.G., Yan Y., De Boer J.M., Baum T.J.,
 Wang X., Husey R.S., Gommers F.J., Henriksat B., Davis E.L.,
 RA Holder J., Schots A., Bakker J.,
 RT "Endogenous cellulases in animals: Isolation of beta-1, 4-
 RT endoglucanase genes from two species of plant-parasitic cyst
 RT nematodes."
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4906-4911(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0P50;
 RA Yan Y., Smant G., Stokkermans J.P.W.G., Wang X., Husey R.S.,
 RA Bakker J., Holder J., Schots A., Davis E.L.;
 RU Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC "1- CATALYTIC ACTIVITY: ENDOPHYLLOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE."
 DR EMBL: AF006052; AAC48327.1;
 DR EMBL: AF052733; AAC15707.1;
 DR HSSP: P07103; 1RGZ
 DR INTERPRO: IPR001547;
 DR PFAM: PF00150; cellulase, 1.
 DR Signal: Hydrolase; glycosidase.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 476 BETA-1,4-ENDOGUCANASE-1.
 SQ SEQUENCE 476 AA; 49815 MW; 2056F90CA0783FF7 CRC64;

Query Match 7.2%; Score 187.5; DB 5; Length 476;
 Best Local Similarity 23.7%; Pred. No. 1.2e-05;
 Matches 120; Conservative 61; Mismatches 211; Indels 115; Gaps 23;

QY 13 LSLILASILFVSG-TSTANANGSGYSGTLLDANGNPFYMRGIN-HGHAWYKD-CAT 68
 DB 13 LELLALCTALYSSLTAVAPFGOLSVGTNLVYGANGQPOLIGSLTFWMOYPOFNAD 72
 QY 69 TAISGANTGANVYR--IVLSDGGQWTKDDIHYRNL-----ISLAEDNHLVAVPEVHDAT 122
 DB 73 TVKALKCKMNNANVIRGAGVDEGGTL--DANTAYNLAVAVIAISGITYIVDMBAHN 130
 QY 123 GYDSIASINRAVDYWIEMRSALIGKEDTVIINANFPG-SWEGDAMADGYKQALPRLN 181
 DB 131 SHP-----DEAVKFFTRTAQ-YGSYPHILYDEPEPSYEW-T-DVLYPIHKVIAALRA 183
 QY 182 AGLNHTLWADAAGGQ-----FPOSIDHYGREVNADFORNTWFSIHMYETAGNASQVR 236
 DB 184 IDKNVILIGPTWSQDVDAVSQNPIDKY-----QNLMTLHFY-----ASSHF 227
 QY 237 TN-----IDRLNDLALVIGFEG--HRTNGDVEDATINSYSEQRGVW-LAMSWKNG 288
 DB 228 TNDLGAKIKTAVNNGLPVFEVTEYGCASGNGNLTDSMSW-----WTLLSLKLSY 280
 QY 289 PEMEYLDLS-----NDWAGNLTAMGNTIYNGPYLRETSRLSTVFPGGSD 335
 DB 281 ANMAISDEKSCALSPTITLANGVSSRWTSNGMVA--YTKKSGVSCSGSSSS 338
 QY 336 GGTSPITLYDEGSMQGTGSSLGCPWAVYEMSSKSHSLKADQLQSLNSOHTLVION 395
 DB 339 SGSSSGSSSGSSSGSSSGSSSG-----SSSGSSSGSSSGSSSGSASISVPSN 390
 QY 396 TSLQONSHIQATYKHAMGSGVNGMTAALYKTHGTYWYSGSPYINGSGTTISLDS 455
 DB 391 T--WNGGGQVNFETKINIGVY-----LGGVVFVSLPSTLT----- 425
 QY 456 NVONLSQVREIGVOFQASDSSGOTS 482
 DB 426 -----GGSMNESAGSGOTSL 441

RESULT 13
 ID 066064 PRELIMINARY; PRT: 481 AA.
 AC 066064;

DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)
 DE ENDOGLUCANASE.
 OC Actinomycetes sp. 40.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-40;
 RA CHO K.K.;
 RU Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U94825; AAC06196.1;
 DR HSSP: O85465; 2A3H
 DR INTERPRO: IPR001547;
 DR INTERPRO: IPR001919;
 DR PFAM: PF00150; cellulase, 1.
 DR PFAM: PF00553; CBD_2; 1.
 DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; UNKNOWN_1.
 SQ SEQUENCE 481 AA; 50750 MW; 50A6A29BEE44057A CRC64;

Query Match 7.2%; Score 187.5; DB 2; Length 481;
 Best Local Similarity 20.6%; Pred. No. 1.2e-05;
 Matches 100; Conservative 76; Mismatches 212; Indels 97; Gaps 21;

QY 38 VSGTLLDANGNPFYMRGIN-HGHAWYKDQAT-TAISGANT-GANTYRIVL-----S 87
 DB 45 ISAPTILDEHGMPPOLNRASHTGLQWGYNKAQFSLRDEMGLNRLVAVYREGGYL 104
 QY 88 DGGQWTKDDITVNLISLADNHLVAVPEV-----DATGYDSIASINRAVDYWI 139
 DB 105 QGSAQMD--KTIOGYOAAATDGLVYIIDWHLVNLNPNNGDATQAES-----FFK 152
 QY 140 MRSALIGKEDTVIINANFPGSWEGDAMADGYKQAL-----PRLNAGNHTLWAD 191
 DB 153 SYAKKYSYGVNIEVCEPRTGT-----PWYDGSNDIYSCTHMAKAIROAGSDAILCG 208
 QY 192 AAGNGQPOSIEDHYGREVNADFORNTWFSIHMYETAGNASQVRTNDRVNDLALVI 251
 DB 209 TWTWS--QDIDAVAGKPLSADGFDNIMYVLF--YATHDDDRAKLQTLNAGTPVFF 263
 QY 252 GEGF--HRTNGDVEDA---NSYSEQRGVGW-LAMSWKNGPEMEY-----LDSNDAG 302
 DB 264 SEFGICDASGGGIDQDSANAMMTLLAHNNISYAMALSNFAETRAFFKPSVATSKWTG 323
 QY 303 NNLAMGNTIYNGPYLRETSRLSTVFPGGSDGTSPTLLYDEGSMQGTG--SSLG 360
 DB 324 DDLTPSAIWLVTYSRKLADPADHAASGTSSSKASASGOTGANGKKSASASSPSAASG 383
 QY 361 GPWAV---TWSKSHSLKADQLQSLNSOHTLVIONTSIQONSRIQATVKHAMGVS 416
 DB 384 GLTFATAIRNOMNGQATYAL--TVSNASGSKH-----EGAMQVTFPDADIAIDI-WG-- 432
 QY 417 GNGMTAALYVTHGTYWYSGSPYINGSGTTISLDSNVONLSQVREIGVOQASDS 476
 DB 433 -----GTIVSHGTHYVAPMDMNTAIEAGASAEIGFN---ASS 468
 QY 477 SGOTS 481
 DB 469 TGOAS 473

RESULT 14
 ID 083012 PRELIMINARY; PRT: 501 AA.
 AC 083012;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)
 DE CELLULOSE.
 OS Bacillus sp.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;

Query Match	7.2%;	Score 187.5;	DB 2;	Length 501;
Best Local Similarity	21.4%;	Pred. No. 1.3e-05;		
Matches 86;	Conservative 72;	Mismatches 140;	Indels 103;	Gaps 19;

RESULT 15

ID	PRELIMINARY;	PRT;	387 AA.
AC	031029;		
DT	01-JAN-1998 (TREMblrel. 05, Created)		
DT	01-JAN-1998 (TREMblrel. 05, Last sequence update)		
DT	01-MAY-2000 (TREMblrel. 13, Last annotation update)		
DE	BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.		
GN	CELA.		
OS	Erwinia carotovora subsp. carotovora.		
OC	Bacteriella; Proteobacteria; gamma subdivision; Enterobacteriaceae.		
OC	Erwinella.		
PN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-1Y34;		
RA	Paik Y.W., Lim S.T., Yun H.D.;		
RL	MOL. Cells 0:0-0(1997).		
DR	EMBL: AF025768; AAC02964.1; -.		
DR	HSSP: 085465; 2A3H.		
DR	INTERPRO: IPR001544; -.		
DR	PFAM: PF00150; cellulase; 1.		

Query Match 7.2%; Score 186.5; DB 2; Length 387;
Best Local Similarity 21.4%; Pred. No. 1.1e-05;
Matches 82; Conservative 63; Mismatches 143; Indels 95; Gaps 16;

QY	6	KKHSITSLILSLASSILFVSGTSTANNNSG-FYVSGTLLDANGNPEYVAGI-NHGHWY	63
Db	10	RKLALGVVTVLWGLSLEFSALSATPYETHQLSTENRVLVDEGGRQLQGLVSGSHGLQWF	69
QY	64	KDQATTAIEGIANTGANTVRIVLSDGQWTRKD-----IHTV	100
Db	70	GD-----YKKDSKRWLPDDMGIVNVRHWMTAADGYISKPLANKV	111
QY	101	RNLISLAEHDHVLAVPEVHDATGDSIASLNRADVITKRSALIGKEDTVIINANWF	160
Db	112	KEAVAAQAQSGVYIIIDWHLISDNEPIYKEQAKTFAEK-AGLYGNSPVIYEIANEPN	170
QY	161	G--SMEDG--AMADGYKQALPRLNAGLNTLTVADAGWGQFPOSITDYGREVFVNDPQR	216
Db	171	GGVWDEIRPYA---LEVETIRSKDPDULIIV---GIGWQDIDH---AADNQLPDP	221
QY	217	NTFESIMYEXAGGNSQVATNIDRYLNDLALVIGFEGHRRHNGD-----VDEATIMSY	271
Db	222	NITVALHF--YAGTHGQFLRITDYAGSRGAALFVSWGISDASGNGPFLPESHMTDF	279
QY	272	SEQSGVWMLAMSKNGNPFWEYEDLS-----NDMAGNNLTANGNTIVNGPYG	318
Db	280	LNNNG-----GSRVNNSLTIDKSEASALAPGASKCGMTEONLSASGKEY-----	324
QY	319	LRERSLSTVFTGGSGDGGSPPT	341
Db	325	RAQIRAAATLGGG--DTPTPT	344

Search completed: December 19, 2000, 16:17:09
Job time: 1010 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:49:31 ; Search time 214.78 Seconds

(without alignments)
47.761 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 MANSGRYVSGTLTYDANGNP.....TIVNGPYGLRSTSTVFT 300

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36:*

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1630	100.0	490	Y54122	Amino acid sequenc
2	1624	99.6	476	Y54123	A mannanase-linker
3	1283	78.7	468	Y44496	Bacillus agaradher
4	1283	78.7	468	Y54125	Amino acid sequenc
5	1283	78.7	493	Y44495	Bacillus agaradher
6	1283	78.7	493	Y54124	Amino acid sequenc
7	1136	69.7	331	Y54127	Amino acid sequenc
8	1075	66.0	369	Y54129	Amino acid sequenc
9	937.5	57.5	320	Y54132	Amino acid sequenc
10	522	32.0	188	Y54133	Amino acid sequenc
11	369	22.6	305	Y54130	Amino acid sequenc
12	316	19.5	132	Y54131	Amino acid sequenc

13	202	12.4	410	18	W12378
14	202	12.4	411	18	W12379
15	202	12.4	411	18	W12381
16	202	12.4	412	18	W12380
17	201	12.3	400	18	W23601
18	201	12.3	400	18	W22521
19	201	12.3	400	19	W57431
20	201	12.3	462	19	W57433
21	200	12.3	409	9	P81843
22	191	11.7	467	17	W05731
23	191	11.7	467	17	W00382
24	174.5	10.7	499	14	R42122
25	171	10.5	352	20	P08472
26	156	9.6	1010	19	W34989
27	152	9.3	800	8	P70420
28	152	9.3	822	13	R26021
29	149.5	9.2	357	16	R77394
30	149.5	9.2	941	11	R07478
31	149.5	9.2	941	16	R77395
32	142	8.7	358	19	W39362
33	142	8.7	521	17	R89827
34	142	8.7	562	21	V69508
35	135	8.3	551	18	W18790
36	127	7.8	472	19	W37243
37	126	7.7	302	19	W43909
38	124	7.6	484	19	W43910
39	118.5	7.3	956	19	W49874
40	118	7.2	476	19	W37241
41	112.5	6.9	360	19	W49870
42	112.5	6.9	386	19	W37242
43	111.5	6.8	360	18	W34565
44	110.5	6.8	532	12	R13329
45	110	6.7	531	16	W01503

ALIGNMENTS

RESULT 1	
Y54122	
ID	Y54122 standard; Protein; 490 AA.
XX	
AC	Y54122;
DT	27-MAR-2000 (first entry)
XX	
DE	Amino acid sequence of a Bacillus mannanase enzyme.
XX	
KW	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW	endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW	galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW	printing paste; plant material degradation; recycled waste paper;
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;
KW	mannan-containing food; coffee extract; cleaning composition;
KW	machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW	contact lens; body-care composition; fabric softener; oil well drilling;
KW	subterranean formation fracture.
XX	
OS	Bacillus sp.
XX	
EH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..30
FT	/note- "signal peptide"
FT	31..330
FT	Domain
FT	/note- "catalytic domain"
FT	Misc-difference 116
FT	/note- "encoded by CTT"
FT	Domain
FT	331..342
FT	/note- "linker"
FT	343..490
FT	/note- "domain of unknown function"
XX	

P300-Ce1B fusion c
P300-Ce1B fusion c
P300-Ce1B fusion c
P300-Ce1B fusion c
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
Sequence of alkali
Cellulase. Bacill
Bacillus cellulase
NK-1 cellulase. B
Actinomycete sp. 3
Terebinthaceae end
Sequence encoded b
Alkaline cellulase
Fragment of alkali
Cellulase. Bacill
Full length Bacill
A. cellulolyticus
A. cellulolyticus
Acidothermus cell
Corrected Bacillus
Globodera rostoch
Globodera glycine
Heterodera glycine
Bankia gouldi glyc
Heterodera glycine
Thermotoga OC1/4V
Thermotoga OC1/4V
Endoglucanase enco
60 kD endoglucanas

PN WO9964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kaupplien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 DR N-PSDB: 245335.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 PI extract, and in cleaning compositions
 XX
 PS Claim 1; Page 208-210; 242pp; English.
 XX
 CC The present sequence represents a mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
 CC of *Bacillus* sp. 1633. The mannanase hydrolyses galactomannans.
 CC Specifically, mannases hydrolyse 1,4-beta-D-mannosidic linkages in
 CC mannan, galactomannan, glucomannan, and galactoglucmannans. The
 CC mannanase protein, or preparations containing it, are used to improve
 CC properties of cellulosic or synthetic fibres, yarn or (non)woven
 CC fabrics (removal of mannan-based sizes or printing pastes). They are
 CC also used to degrade or modify plant materials (particularly recycled
 CC waste paper, paper making pulp, or material containing guar or locust
 CC bean gums (thickeners), or to reduce viscosity of mannan-containing
 CC foods (or feeds). The mannases are also used to process coffee
 CC extracts (to inhibit gel formation); in cleaning compositions (for
 CC machine washing of fabrics, as hard-surface cleaners, for hand or
 CC machine dishwashing, also in oral, dental, contact lens or body-care
 CC compositions) where they remove mannan-containing soils and prevent
 CC binding of some soils to cellulosics; and in fabric softeners. They
 CC can also be used in oil well drilling to fracture subterranean
 CC formations.
 CC
 CC Sequence 490 AA.
 SO

Query Match 100.0%; Score 1630; DB 21; Length 490;
 Best Local Similarity 100.0%; Pred. No. 7.8e-141;
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NANSFYVSGTLYDANGPFYRGINGHANYKDOATTAIGANTGANTVRIYLSGCG 60
 DB 31 nansfyvsgtlydangpfrmginhanykdattatgiantgntvriylsdg 90
 QY 61 QWTKDHTVRRLISLAEDNHLVAVPEVDATGYDSIASLANAVYVTEMRALGKEVT 120
 DB 91 qwtkdhtvrllislaednhlvapevdatgydsiaslanavyvtemrallgkect 150
 QY 121 VLIINANEFSGWEDAMADGYKQAIPLRLNAGLWHTLWDAAGWQFOSIHDYGREVF 180
 DB 151 vliinanevsgwedamadykqalprlnaglwhtlwnvdaagwqfpgsihdygrevf 210
 QY 161 NADPQRTWFSIHMTETVAGGNASQVFTNIDRVLNDLALVIGFGRHRTNGDVDEATWS 240

DB 211 naddpqrntwfsihmetyaggnasqvftndrvnlndlalvigfgrhrtngdvdeatws 270
 QY 241 YSEQRGVGTAWSMKNGPWEWEYLDSDMAGNNLTJANGNTIVNGPYGLRETSRLSTVFT 300
 DB 271 yseqrvgvtawsmkngpweyldsdmagnnltjangntivngpyglretsrilstvft 330
 RESULT 2
 Y54123
 ID Y54123 standard; Protein; 476 AA.
 XX
 AC Y54123;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE A mannanase-linker-cellulose binding domain fusion protein.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; *Bacillus* sp. 1633; galactomannan;
 KW 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
 KW galactoglucmannan; cellulosic fibre; synthetic fibre; yarn; fabric;
 KW printing paste; plant material degradation; recycled waste paper;
 KW paper making pulp; guar; locust bean gum; thickener; viscosity;
 KW mannan-containing food; coffee extract; cleaning composition;
 KW machine washing; hard-surface cleaner; dishwashing; oral; dental;
 KW contact lens; body-care composition; fabric softener; oil well drilling;
 KW subterranean formation fracture; cellulose binding domain.
 KW
 OS Synthetic.
 OS *Bacillus* sp.
 OS *Clostridium thermocellum*.
 XX
 PX WO9964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kaupplien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 DR N-PSDB: 245335;
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 PI extract, and in cleaning compositions
 XX
 PS Example 4; Page 211-212; 242pp; English.
 XX
 CC The present sequence represents a mannanase-linker-cellulose binding
 CC domain fusion protein. Mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
 CC hydrolyses galactomannans. Specifically, mannases hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannan, galactomannans, glucomannans,
 CC and galactoglucmannans. The mannanase protein, or preparations
 CC containing it, are used to improve properties of cellulosic or

CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC celluloses; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.

SQ Sequence 476 AA;

Query Match 99.6%; Score 1624; DB 21; Length 476;
Best Local Similarity 100.0%; Pred. No. 2,6e-140;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ANSGFTYSGITLTDANGNPFYMRGINHGHANTKDOATTALIGIANTGANTVRYVLSGQ 61
DB 1 ansqfyvsgtlltydangnpfymrglnbhavkdkqatlaeglanqantvryvlsdgq 60
QY 62 WTKDDIHTVRNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWIEMRSALIGKEDTV 121
DB 61 wtkddihvrrnlislaednhlvavpevhdacgysiaslnravdywlemrallgkedtv 120
QY 122 IINTANEMFGSMEDGAWMDGYKQAIPLRLNAGLNTLTVDAAGWQFQSIHDYGREYFN 181
DB 121 iintanewfgswgdawmdgdykqalprlrnaglnhtlmvdaagwqfqsldygreyn 180
QY 182 ADPQNTWFSIHMTETVAGNAGSOVRTNIDRYLNODLALVIGFGRHRTNGVDVATIMSY 241
DB 181 adpqnwmfshmtetvagnagsovrtndrylnodlalvigfgrhrtngvdeatimsy 240
QY 242 SEQGVGWLANSWKGPEWEYELDSNDMAGNNTLVNGNTVNGPYGIRERSLTSTVFT 300
DB 241 seqgtvwlanswkgpeweyelidsnwmagnntlvngntlvngpygltresltstvt 299

RESULT 3
ID Y44496 standard; Protein; 468 AA.
AC Y44496;
XX 27-MAR-2000 (first entry)
DE Bacillus agaradhaerens Clone MB594, Mannanase enzyme.
XX
KW Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KM detergent composition; mild-branched anionic surfactant; washing;
XX cosmetic stain; food stain.
OS
XX Bacillus agaradhaerens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Peptide 1..31
FT /label= Signal_peptide
FT Protein 32..468
FT /label= Mature_Mannanase
XX
XX MO9964552-A1.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1998; 98MO-US12026.
XX
XX 10-JUN-1998; 98MO-US12026.
XX
XX 10-JUN-1998; 98MO-US12026.
XX
XX (PROC) PROCTER & GAMBLE CO.

XX Bettli JP, Thoen CAUK;
XX
XX WPI: 2000-116536/10.
DR N-PSDB; 229846;
XX
XX Detergent composition for removing greasy stains such as cosmetics,
PT food stains and body soils -
XX
XX Disclosure; Page 103-104; 113pp; English.
XX
XX The present sequence is the B. agaradhaerens clone MB594, alkaline
CC mannanase enzyme. Clone MB594 is derived from B. agaradhaerens strain
CC NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
CC from 7.5-10.5. It can be used in a detergent composition along with a
CC mild-branched anionic surfactant. The detergent composition may be used
CC for washing purposes, soaking/pre-treatment of stained fabric, hard
CC surface cleaning and for removal of cosmetic and/or food stains. This
CC composition provides excellent cleaning effect at low temperature.

SQ Sequence 468 AA;

Query Match 78.7%; Score 1283; DB 21; Length 468;
Best Local Similarity 77.3%; Pred. No. 3.5e-109;
Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

QY 2 ANSGFTYSGITLTDANGNPFYMRGINHGHANTKDOATTALIGIANTGANTVRYVLSGQ 61
DB 32 astfydgntlltydangnpfymrglnbhavkktasalapaleqantlrvlsdgq 91
QY 62 WTKDDIHTVRNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWIEMRSALIGKEDTV 121
DB 92 wtkddihvrrnlislaednhlvavpevhdacgysiaslnravdywlemkdallgkedtv 151
QY 122 IINTANEMFGSMEDGAWMDGYKQAIPLRLNAGLNTLTVDAAGWQFQSIHDYGREYFN 181
DB 152 iintanewfgswgdawmdgdykqalprlrnaglnhtlmvdaagwqfqsldygreyn 211
QY 182 ADPQNTWFSIHMTETVAGNAGSOVRTNIDRYLNODLALVIGFGRHRTNGVDVATIMSY 241
DB 212 adpqnwmfshmtetvagnagsovrtndrylnodlalvigfgrhrtngvdeatimsy 271
QY 242 SEQGVGWLANSWKGPEWEYELDSNDMAGNNTLVNGNTVNGPYGIRERSLTSTVFT 300
DB 272 seqgtvwlanswkgpeweyelidsnwmagnntlvngntlvngpygltresltstvt 330

RESULT 4
ID Y54125 standard; Protein; 468 AA.
AC Y54125;
XX
XX 27-MAR-2000 (first entry)
DE Amino acid sequence of a Bacillus mannanase enzyme.
XX
XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; galactomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture.
XX
XX
XX Synthetic.
OS
OS Bacillus agaradhaerens.
XX
XX MO9964619-A2.

PD 16-DEC-1999. 99WO-DK00314.
 XX 10-JUN-1999;
 PR 10-JUN-1998; 9805-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 9805-0105970.
 PR 28-OCT-1998; 9805-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 9905-0123543.
 PR 10-MAR-1999; 9905-0123623.
 PR 10-MAR-1999; 9905-0123641.
 PR 11-MAR-1999; 9905-0123642.
 XX (NOVO) NOVO-NORDISK AS.
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 DR WPI: 2000-105891/09.
 DR N-PSDB; 245338.
 XX New mannanses for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions -
 PS Example 5; Page 215-216; 242pp; English.
 XX The present sequence represents a mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The present mannanase is a synthetic variant of the mannanase of
 CC Y54124, in which the C-terminus of the protein was changed due to
 CC design of a lower PCR primer used for amplification. The mannanase
 CC hydrolyses galactomannans. Specifically, mannanses hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, or
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC slimes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulp,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 CC Sequence 468 AA;
 SO

Query Match 78.7%; Score 1283; DB 21; Length 468;
 Best Local Similarity 77.3%; Pred. No. 3.5e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

OY 2 ANSGFVSGTTLTDANGNPFVNRGINHGAWYKQQTATIEGANTGATVRIYLSGGO 61
 DB 32 asgfyvsgntlydangpfrvnrghnawykdtastalpaegaqantlrvlsdgg 91
 OY 62 WTKDDIHTVRNLTSLAEDNHLVAVPEVDATGYDSIASLRAVDYWIEMRSALIGKEDTV 121
 DB 92 wekdidditrevlelaeqhmvavvevdatsrdsdinravdywlemkdaligkredtv 151
 OY 122 IINIAENMGSWEGDAGAGYKQAIPTLRNAGINFTLMDAGWGQFPOSIDHYGESEVN 181
 DB 152 Ilniaenwyswgdagagydaiptlrdagltchlmvdaagwsgypspidhygeevdn 211
 OY 182 ADPQRTMFSIMHYEYAGNASQVTRINDRVNLQDLALVIGERGHRTNGDVDEATIMSY 241

DB 212 adplixntmsimhyeyagdgantvrsndrvldqalavlgfgrhndgdtdedtlisy 271
 OY 242 SEQRGVGLAWGSKNGPMEYTLDSNDWAGNNLTWAGNTIVNGPYGLRETSRLSTVFT 300
 DB 272 seetgrylawwkwngstewdyldtsedwagqltdwgnrvlvgadgylqetkpsvtft 330

RESULT 5
 Y44495
 ID Y44495 standard; Protein: 493 AA.
 XX Y44495;
 AC 27-MAR-2000 (first entry)
 XX
 DT Bacillus agaradherens NCIMB 40482, Mannanase enzyme.
 XX
 DE Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
 KW detergent composition; mid-branched anionic surfactant; washing;
 KW cosmetic stain; food stain.
 XX
 OS Bacillus agaradherens.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..32
 FT Protein /label- Signal_peptide
 FT 33..343
 FT /label- Mature_Mannanase
 XX
 XX W0964552-AI.
 PD 16-DEC-1999.
 XX 10-JUN-1998; 98WO-US12026.
 PR 10-JUN-1998; 98WO-US12026.
 PR 10-JUN-1998; 98WO-US12026.
 XX (PROC) PROCTER & GAMBLE CO.
 PA Bettiol JP, Thoen CAUK;
 PI Detergent composition for removing greasy stains such as cosmetics,
 DR food stains and body soils -
 DR N-PSDB; 229845.
 XX WPI: 2000-116536/10.
 DR Detergent composition for removing greasy stains such as cosmetics,
 PT food stains and body soils -
 PS Disclosure; Page 102; 113pp; English.
 XX The present sequence is the B. agaradherens NCIMB 40482, alkaline
 CC mannanase enzyme. Mannanase shows maximum activity at pH ranging
 CC from 7.5-10.5. It can be used in a detergent composition along with a
 CC mid-branched anionic surfactant. The detergent composition may be used
 CC for washing purposes, soaking/pre-treatment of stained fabric, hard
 CC surface cleaning and for removal of cosmetic and/or food stains. This
 CC composition provides excellent cleaning effect at low temperature.
 CC Sequence 493 AA;
 SO

Query Match 78.7%; Score 1283; DB 21; Length 493;
 Best Local Similarity 77.3%; Pred. No. 3.7e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

OY 2 ANSGFVSGTTLTDANGNPFVNRGINHGAWYKQQTATIEGANTGATVRIYLSGGO 61
 DB 32 asgfyvsgntlydangpfrvnrghnawykdtastalpaegaqantlrvlsdgg 91
 OY 62 WTKDDIHTVRNLTSLAEDNHLVAVPEVDATGYDSIASLRAVDYWIEMRSALIGKEDTV 121
 DB 92 wekdidditrevlelaeqhmvavvevdatsrdsdinravdywlemkdaligkredtv 151

QY 122 IINIANEFGSWEGDAMADYKQAIPLRLNAGLNTLWDAWGQFPOSTIDYGREYFN 181
 DB 132 IINIANEFGSWEGDAMADYKQAIPLRLNAGLNTLWDAWGQFPOSTIDYGREYFN 211
 QY 182 ADPQRTMFSIMMYEYAGGNSQVNTNDIRVYNODLALVIGFGRHNTGVDDEATIMSY 241
 DB 212 adpLxntmfsImyeyaggdantvrsndrvidqdaIvIgefgnrhndgdvdedtllsy 271
 QY 242 SEQRGVGLWAMSGKNGPEWEXYLDISNDWAGNNLWAMGNTIYNGPYGLRETSRLSTVFT 300
 DB 272 seetgtgylawswkngstewdyldIsedwagqhlctwgnrIvHgagdgIqetskpsvtvf 330
 RESULT 6
 Y54124 standard; Protein; 493 AA.
 Y54124;
 27-MAR-2000 (first entry)
 Amino acid sequence of a *Bacillus* mannanase enzyme.
 Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 mannan; galactomannan; glucomannan; galactoglucomannan; cellulose
 synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 plant material degradation; recycled waste paper; paper making pulp;
 guar; locust bean gum; mannan-containing food; coffee extract;
 cleaning composition; machine washing; hard-surface cleaner;
 dishwashing; oral; dental; contact lens; body-care composition;
 fabric softener; oil well drilling; subterranean formation fracture.
Bacillus agaradhaerens.
 Key Location/Qualifiers
 Peptide 1..31
 /note= "signal peptide"
 Domain 32..344
 /note= "catalytic domain"
 Domain 345..493
 /note= "domain of unknown function"
 W09964619-A2.
 16-DEC-1999.
 10-JUN-1999; 99WO-DK00314.
 10-JUN-1998; 98US-0111256.
 20-OCT-1998; 98DK-0001340.
 20-OCT-1998; 98DK-0001341.
 28-OCT-1998; 98US-0105970.
 28-OCT-1998; 98US-0106054.
 23-DEC-1998; 98DK-0001725.
 05-MAR-1999; 99DK-0000306.
 05-MAR-1999; 99DK-0000307.
 05-MAR-1999; 99DK-0000308.
 05-MAR-1999; 99DK-0000309.
 09-MAR-1999; 99US-0123543.
 10-MAR-1999; 99US-0123623.
 10-MAR-1999; 99US-0123641.
 11-MAR-1999; 99US-0123642.
 (NOVO) NOVO-NORDISK AS.
 Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 WPI; 2000-105891/09.
 N-PSDB; 245337.
 New mannanases for treatment of textiles, plant material and coffee
 extract, and in cleaning compositions

XX Claim 35; Page 213-214; 242pp; English.
 PS The present sequence represents a mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The
 CC mannanase hydrolyses galactomannans. Specifically, mannanases hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
 CC and galactoglucomannans. The mannanase protein, or preparations
 CC containing it, are used to improve properties of cellulosic or
 CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics), as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 SQ Sequence 493 AA;
 Query Match 78.7%; Score 1283; DB 21; Length 493;
 Best local similarity 77.3%; Pred. No. 3.7e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;
 QY 2 ANSGFYVSGTTLTYDANGNPFVWRGINHGHWKQDQATAEIGANTGATVRIYSDGQ 61
 DB 32 estgtgvtgnltlydangpftmrglnhghwykdaaIpaIaeqgnaltvIsdgg 91
 QY 62 WTKDDIHTVRNLISLAEEDNHLVAPEVDATGYSIASINRAVDYIEMRSALICKEDTV 121
 DB 92 wekdldtIreIvIeIaekmwaVevendatgrdsrdlnravdvIemkdlIsKedtv 151
 QY 122 IINIANEFGSWEGDAMADYKQAIPLRLNAGLNTLWDAWGQFPOSTIDYGREYFN 181
 DB 152 IINIANEFGSWEGDAMADYKQAIPLRLNAGLNTLWDAWGQFPOSTIDYGREYFN 211
 QY 182 ADPQRTMFSIMMYEYAGGNSQVNTNDIRVYNODLALVIGFGRHNTGVDDEATIMSY 241
 DB 212 adpLxntmfsImyeyaggdantvrsndrvidqdaIvIgefgnrhndgdvdedtllsy 271
 QY 242 SEQRGVGLWAMSGKNGPEWEXYLDISNDWAGNNLWAMGNTIYNGPYGLRETSRLSTVFT 300
 DB 272 seetgtgylawswkngstewdyldIsedwagqhlctwgnrIvHgagdgIqetskpsvtvf 330
 RESULT 7
 Y54127 standard; Protein; 331 AA.
 Y54127;
 27-MAR-2000 (first entry)
 Amino acid sequence of a *Bacillus* mannanase enzyme.
 Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 mannan; galactomannan; glucomannan; galactoglucomannan; cellulose
 synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 plant material degradation; recycled waste paper; paper making pulp;
 guar; locust bean gum; mannan-containing food; coffee extract;
 cleaning composition; machine washing; hard-surface cleaner;
 dishwashing; oral; dental; contact lens; body-care composition;
 fabric softener; oil well drilling; subterranean formation fracture.
Bacillus sp.
 Key Location/Qualifiers

FT	Peptide	1..32	/note="signal peptide"
FT	Domain	33..331	/note="catalytic domain"
FT	Domain		
FM	W09964619-A2.		
PD	16-DEC-1999.		
XX			
PF	10-JUN-1999;	99WO-DK00314.	
XX			
PR	10-JUN-1998;	98US-0111256.	
PR	20-OCT-1998;	98DK-0001340.	
PR	20-OCT-1998;	98DK-0001341.	
PR	28-OCT-1998;	98US-0105970.	
PR	28-OCT-1998;	98US-0106054.	
PR	23-DEC-1998;	98DK-0001725.	
PR	05-MAR-1999;	99DK-0000306.	
PR	05-MAR-1999;	99DK-0000307.	
PR	05-MAR-1999;	99DK-0000308.	
PR	09-MAR-1999;	99DK-0000309.	
PR	10-MAR-1999;	99US-0123543.	
PR	10-MAR-1999;	99US-0123623.	
PR	11-MAR-1999;	99US-0123641.	
PR	11-MAR-1999;	99US-0123642.	
PA	(NOVO) NOVO-NORDISK AS.		
PI	Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;		
PI	WPI; 2000-105891/09.		
DR	N-PsDB; Z45340.		
XX			
PT	New mannanses for treatment of textiles, plant material and coffee		
PT	extract, and in cleaning compositions		
XX			
PS	Claim 35; Page 220-221; 242pp; English.		
XX			
CC	The present sequence represents a <i>Bacillus mannanase</i> (also known as		
CC	mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).		
CC	The mannase hydrolyses galactomannans. Specifically, mannases		
CC	hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,		
CC	glucomannans, and galactoglucomannans. The mannase protein, or		
CC	preparations containing it, are used to improve properties of cellulosic		
CC	or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based		
CC	slies or printing pastes). They are also used to degrade or modify		
CC	plant materials (particularly recycled waste paper, paper making pulps,		
CC	or material containing guar or locust bean gums (thickeners), or to		
CC	reduce viscosity of mannan-containing foods or feeds). The mannases		
CC	are also used to process coffee extracts (to inhibit gel formation); in		
CC	cleaning compositions (for machine washing of fabrics, as hard-surface		
CC	cleaners, for hard or machine dishwashing, also in oral, dental, contact		
CC	lens or body-care compositions) where they remove mannan-containing		
CC	soils and prevent binding of some soils to celluloses; and in fabric		
CC	softeners. They can also be used in oil well drilling to fracture		
CC	subterranean formations.		
XX			
SQ	Sequence 331 AA;		
XX			
XX			
Query Match	69.7%;	Score 1136;	DB 21; Length 331;
Best Local Similarity	68.7%;	Pred. No. 5.7e-96;	
Matches 204; Conservative 47;	Mismatches 46;	Indels 0;	Gaps 0
3	NSGFVSGTGVTDANGNPVVRMGINNGHMYDQATATAGTANSGANVRYTSSGGOM	62	
Db	34 hsgfivngtllvdangnpvrmgimngshwrtfgeletsmrglsqgancitrvlslngqrw	93	
63	TKDIIHTVKNLISLAEDNHLVAPEVHADATGYDSIASLNKRAVDYTEMKSALIGKEDPTV	122	
Db	94 qdddnmwasvsiaseghmqiaavlervhdatsgmfnfdgaavdyvtemkdvlvgkcdv	153	
123	INIAENWGSWEGMADADYKQAIPIRLNAGNLHNTLVADNAGGCPFQSIHDYGRGVFN	182	

Db	154	linineyrgawdgawarqyqnaqlrlrlnaglshtclmvaagaygqypgavdyggevina	213
Oy	183	DPOSNTFSTHMYETACGNASQVRYTINDRYLNDOLAVIGEGFHRRTNGDVDEATIMSY	242
Db	214	dprntfsvlmuyeyagsgdantvrrldslslsgqlavlgfghwlydgdvadtllsy	273
Oy	243	EORVGWLANSWKNGPEWEXELDSMDAGNNLTAMGNTIVNGPGALRETSLSLTFE	299
Db	274	qgrivgwlaawswngsegveylslsdnfdgrrltwgdrlvngpqlrqlskrsyvf	330
RESULT	8		
ID	Y54129	standard; Protein; 369 AA.	
XX	Y54129		
AC	Y54129		
XX	Y54129		
DT	27-MAR-2000	(first entry)	
XX			
DE		Amino acid sequence of a <i>Bacillus</i> sp. AA349 mannanase enzyme.	
XX			
KW		Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;	
KW		endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;	
KW		mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre	
KW		synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;	
KW		plant material degradation; recycled waste paper; paper making pulp;	
KW		guar; locust bean gum; mannan-containing food; coffee extract;	
KW		cleaning composition; machine washing; hard-surface cleaner;	
KW		dishwashing; oral; dental; contact lens; body-care composition;	
KW		fabric softener; oil well drilling; subterranean formation fracture.	
XX			
XX		<i>Bacillus</i> sp.	
XX			
XX		Key	
FT		Location/Qualifiers	
FT		Peptide	1..25
FT		/note="signal peptide"	
FT		Domain	26..67
FT		/note="N-terminal domain of unknown function"	
FT		Domain	68..369
FT		/note="catalytic domain"	
XX			
XX		WO9964619-A2.	
XX			
XX		16-DEC-1999.	
XX			
XX		10-JUN-1999;	99WO-DK00314.
XX			
XX		10-JUN-1998;	98US-0111256.
XX		20-OCT-1998;	98DK-0001340.
XX		20-OCT-1998;	98DK-0001341.
XX		28-OCT-1998;	98US-0105870.
XX		28-OCT-1998;	98US-0106054.
XX		23-DEC-1998;	98DK-0001725.
XX		05-MAR-1999;	99DK-0000306.
XX		05-MAR-1999;	99DK-0000307.
XX		05-MAR-1999;	99DK-0000308.
XX		05-MAR-1999;	99DK-0000309.
XX		03-MAR-1999;	99US-0123543.
XX		10-MAR-1999;	99US-0123623.
XX		10-MAR-1999;	99US-0123641.
XX		11-MAR-1999;	99US-0123642.
XX			
PA		(NOVO) NOVO-NORDISK AS.	
XX			
XX		Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;	
XX		WPI; 2000-105891/09.	
DR		N-PSDB; Z45342.	
XX			
XX		New mannanases for treatment of textiles, plant material and coffee	
PT		extract, and in cleaning compositions	
XX			

Db 147 vlinanewyawsrsvaagayagalpirlsaglahlilvdaagwgypaslhgeradvf 206
 Oy 181 NADPQRTMFTSHMYEYAGNAGSVYRINIDRYLNDLALVIGFEGHRTNGDPVDEATIMS 240
 Db 207 asdpkntcmfshmyeyagadatrsvsenldgylaenlavigefghrhdgdvdedalla 266
 Oy 241 YSEQRGVMGLAWSKNGKNGPEWETDLSDMAGNNITLNGNTIVNCPYGLRETSRL 295
 Db 267 ylaerwygylawswygsnagvgyelidtegpsg-pltswgerlvygmglykyladh1 320

RESULT 10
 ID Y54133 standard; Protein; 188 AA.
 AC Y54133;
 DT 27-MAR-2000 (first entry)

XX Amino acid sequence of a partial Bacillus sp. mannanase enzyme.
 DE
 XX
 XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 XX
 XX
 OS Bacillus sp.
 XX
 PN WO964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 XX 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 98US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 XX (NOVO) NOVO-NORDISK AS.
 PA
 XX
 XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 PI WPI; 2000-105891/09.
 DR N-PSDB; 245346.
 DR
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 extract, and in cleaning compositions
 PS Claim 35; Page 229-230; 242pp; English.
 XX
 XX The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based

CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 CC
 XX
 SO Sequence 188 AA;
 S0

Query Match 32.0%; Score 522; DB 21; Length 188;
 Best Local Similarity 59.9%; Pred. No. 3e-40;
 Matches 97; Conservative 20; Mismatches 45; Indels 0; Gaps 0;

Oy 1 NANSGETTYCTVDANCPVFMGNGHAWYKQATFATEGIANTGANTVRIVLSDG 60
 Db 27 saqsfihvkgfelleldkngdpyrmgvnhgshwtkqdeelpaiaetganvrvivlsng 86
 Oy 61 QWTKDITVNTLSIAEDNHLVAPVEYDAGYDSIASLNRAVDYWTENRSALIGKEDT 120
 Db 87 qwekdaseelarvlatetyglttvlevhdatsgdndpddlkavdyvlemadvlykgtedr 146
 Oy 121 VIINTANFEGSWEGDAMADGKQALPRLNAGLNTMTMDA 162
 Db 147 vlinanewyawsrsvdhwakayagalpirlsaglahlilvda 188

RESULT 11
 ID Y54130 standard; Protein; 305 AA.
 AC Y54130;
 DT 27-MAR-2000 (first entry)

XX Amino acid sequence of a partial Bacillus sp. mannanase enzyme.
 DE
 XX
 XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 XX
 XX
 OS Bacillus sp.
 XX
 PN WO964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 XX 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 98US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.


```

XX AC W12378;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 1 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. N4 (ATCC 21833).
XX FH Key
XX FT 1.26 Location/Qualifiers
XX FT /label= Sig.peptide
XX FT /note= "Hybrid between P300 (aa1-12) and
XX FT CelB (aa13-26) signal peptides"
XX FT Protein
XX FT 27..410
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN MO9706181-A1.
XX PD 20-FEB-1997.
XX PF 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENK ) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB; T63226.
XX FT System for increased expression of cellulase and xylanase in
XX FT Bacillus - contains gene under control of elements from B.
XX FT licheniformis alkaline protease gene
XX PS Disclosure: Fig 7,7A,7B; 37pp; English.
XX CC The polypeptide product (W12378) of P300-CelB fusion construct 1
XX CC (T63226) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also
XX CC W12379-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 410 AA;

Query Match 12.4%; Score 202; DB 18; Length 410;
Best Local Similarity 25.0%; Pred. No. 1.4e-10;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

OY 8 VSGTLYDANGNFVARGI-NHGAWYKDOATTAIGIANFGANRYRIYLSDGQGWTKDD 66
DB 41 lmgelavndrgpvtlkgmshglqygy-----qfnyesamwlridd 82
OY 67 -----IHT-----VRNLSIAEDNHLVWPEVHDANGYSISLNA 103
DB 83 wgtlviraamytesggyledpsvkekykeaveaalglgylldwlllednprlykeaa 142
OY 104 VDWIMRSLILIKEDPTVIINIANEWFGS---WEGD--AMADGYKQAIPLRNAGLHTL 158
DB 143 kdfidam-selygdynpviyleanepngsdvtdnqtkpyae---evipvlrindpnll 198

```

```

OY 159 KYDAAGMGQFPOSIHDIKREVFNADPQNTMESTHMTAYAGNSOYRINDRYLNODLA 218
DB 199 lv---gtgtwsqdvh-headnqltdp--nvmayathf--yagthgnrlrdyvdalddgaa 250
OY 219 LVIGEGHRTNGD-----VDPATI-MSYSEORGWGLWMS 253
DB 251 lrvsewgtseatgdgvtldeagvldfmdernlswanws 290

RESULT 14
W12379
ID W12379 standard; Protein; 411 AA.
XX AC W12379;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 2 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. N4 (ATCC 21833).
XX FH Key
XX FT 1.26 Location/Qualifiers
XX FT /label= Sig.peptide
XX FT /note= "Hybrid between P300 (aa1-5) and
XX FT CelB (aa8-26) signal peptides"
XX FT Protein
XX FT 27..411
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN MO9706181-A1.
XX PD 20-FEB-1997.
XX PF 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENK ) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB; T63227.
XX FT System for increased expression of cellulase and xylanase in
XX FT Bacillus - contains gene under control of elements from B.
XX FT licheniformis alkaline protease gene
XX PS Disclosure: Fig 8,8A,8B; 37pp; English.
XX CC The polypeptide product (W12379) of P300-CelB fusion construct 2
XX CC (T63227) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also W12378,
XX CC W12380-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 411 AA;

Query Match 12.4%; Score 202; DB 18; Length 411;
Best Local Similarity 25.0%; Pred. No. 1.4e-10;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

```


QY 8 VSGTLLYDANGNPFVNRGI-NHGHANYKDOATTAIEGINTANTGANYRYILSDGGOWTKDD 66
 DB 42 IANGELVNDIRGEVPVQIKGMSHGLQVY-----GFVYESMKWLRDD 83
 QY 67 -----IHT-----VRNLISLAEDNHLVAVPEVDATGYDSIASLNR 103
 DB 84 WGLTVFIRAAMYSSGGYLEDPSVKEKVEAEALDGLYVLDVHLLSDNDPNLYKEEA 143
 QY 104 VDWIEMRSALIGKEDTVIINIANEWFSGS---WEGD--AMADGYKQAIPLRNAGLNHTL 158
 DB 144 KdIFDEM-selygdyPNVlyelanepgsdvtwdnqIKpyae---evlPvIrndpn11 199
 QY 159 MYDAGMGQFPOSIDHYGREVFENADPQRTMFSIHMYEYAGNASQVFRNIDRVLNODIA 218
 DB 200 IV---gtgIwsgdvH-haadnqITdp--nwmYafhf--YagthgnlrqdvYaldqgaa 251
 QY 219 LVIGERGHRTNGD---VDEATI-MSYSEORGVGLWAS 253
 DB 252 IIVSEWGTSEATGdggvfldeeqvldfndernlswaans 291

RESULT 15
 W12381
 ID W12381 standard; Protein: 411 AA.
 AC W12381;
 DT 17-JUN-1997 (first entry)
 DE P300-CelB fusion construct 4 polypeptide product.
 XX
 KW Cellulase; xylanase; alkaline protease; P300; CelB.
 OS Chimeric Bacillus licheniformis ATCC 53926;
 OS Chimeric Bacillus licheniformis;
 OS Chimeric Bacillus sp. N4 (ATCC 21833).
 XX
 FT Key Location/Qualifiers
 FT Peptide 1..27
 FT /label= Sig_peptide
 FT /note= "hybrid between P300 (aa1-22) and
 FT Protein 28..411
 FT /label= Mat_protein
 FT /note= "mature CelB cellulase"
 FT
 FT
 XX W09706181-A1.
 PN
 XX
 PD 20-FEB-1997.
 XX
 PF 08-AUG-1996; 96W0-US12545.
 XX
 PR 07-AUG-1996; 96US-0694346.
 PR 10-AUG-1995; 95US-0002106.
 XX
 PA (HENK) HENKEL CORP.
 XX
 PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
 DR WPI: 1997-154208/14.
 DR N-PSDB; T63229.
 XX
 PT System for increased expression of cellulase and xylanase in
 PT Bacillus - contains gene under control of elements from B.
 PT licheniformis alkaline protease gene
 XX
 PS Disclosure: Fig 10,10A,10B; 37pp; English.
 XX
 CC The polypeptide product (W12379) of P300-CelB fusion construct 2
 CC (T63227) comprises a hybrid signal peptide, formed between the
 CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
 CC alkaline protease and Bacillus licheniformis ATCC 53926 (P300)
 CC and the mature CelB alkalophilic cellulase of Bacillus sp. N4.

CC Expression in Bacillus sp. host cells of P300-CelB fusion
 CC constructs 1-4 (see also W12378-80) provides a 10-40 fold
 CC improvement in prodn. of the alkalophilic cellulase in comparison
 CC to expression of the native gene, and an earlier start to the
 CC prodn. of enzyme in the fermentation process.
 XX
 SQ Sequence 411 AA;

Query Match 12.4%; Score 202; DB 18; Length 411;
 Best Local Similarity 25.0%; Pred. No. 1,4e-10;
 Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

QY 8 VSGTLLYDANGNPFVNRGI-NHGHANYKDOATTAIEGINTANTGANYRYILSDGGOWTKDD 66
 DB 42 IANGELVNDIRGEVPVQIKGMSHGLQVY-----GFVYESMKWLRDD 83
 QY 67 -----IHT-----VRNLISLAEDNHLVAVPEVDATGYDSIASLNR 103
 DB 84 WGLTVFIRAAMYSSGGYLEDPSVKEKVEAEALDGLYVLDVHLLSDNDPNLYKEEA 143
 QY 104 VDWIEMRSALIGKEDTVIINIANEWFSGS---WEGD--AMADGYKQAIPLRNAGLNHTL 158
 DB 144 KdIFDEM-selygdyPNVlyelanepgsdvtwdnqIKpyae---evlPvIrndpn11 199
 QY 159 MYDAGMGQFPOSIDHYGREVFENADPQRTMFSIHMYEYAGNASQVFRNIDRVLNODIA 218
 DB 200 IV---gtgIwsgdvH-haadnqITdp--nwmYafhf--YagthgnlrqdvYaldqgaa 251
 QY 219 LVIGERGHRTNGD---VDEATI-MSYSEORGVGLWAS 253
 DB 252 IIVSEWGTSEATGdggvfldeeqvldfndernlswaans 291

Search completed: December 19, 2000, 16:49:32
 Job time: 1182 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:02:41 ; Search time 156.76 Seconds
(without alignments)
32.077 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630

Sequence: 1 NANSGRVYSCTILYDANGNP.....TIVNGPYGLREMSRLSTVPT 300

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA: *
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	12.3	400	2	US-08-713-298B-2
2	201	12.3	400	2	US-08-870-180B-2
3	201	12.3	400	3	US-08-814-052-4
4	201	12.3	400	3	US-08-812-829-4
5	201	12.3	462	2	US-08-870-180B-13
6	191	11.7	467	2	US-08-727-548-2
7	171.5	10.5	430	2	US-08-924-440-2
8	142	8.7	358	1	US-08-604-913B-11
9	142	8.7	521	1	US-08-276-213-3
10	135	8.3	531	2	US-08-033-537A-1
11	110	6.7	531	2	US-07-862-388B-7
12	100	6.1	429	1	US-08-745-977-4
13	100	6.1	429	1	US-09-040-699A-4
14	99.5	6.1	317	2	US-09-066-075-2
15	99.5	6.1	317	2	US-08-615A-2
16	99.5	6.1	317	3	US-08-951-889-2
17	97.5	6.0	374	3	US-08-733-433-1
18	95.5	5.9	735	2	US-08-313-185-8
19	95.5	5.9	735	2	US-08-459-499-9
20	94	5.8	398	2	US-08-853-659A-45
21	93.5	5.7	566	2	US-07-863-588B-4
22	90.5	5.6	735	2	US-08-459-489-12
23	88	5.4	531	1	US-08-531-601-1
24	88	5.4	531	2	US-08-859-032-1
25	88	5.4	535	1	US-08-737-597-10
26	86	5.3	461	1	US-08-672-571A-3
27	86	5.3	490	1	US-08-672-571A-1
28	85.5	5.2	385	2	US-08-387-942C-26

29	85.5	5.2	997	2	US-08-387-942C-4	Sequence 4, Appl
30	85.5	5.2	1248	2	US-08-348-353-17	Sequence 17, Appl
31	85.5	5.2	1248	2	US-08-465-965-17	Sequence 17, Appl
32	85.5	5.2	1248	3	US-08-465-966-17	Sequence 17, Appl
33	85	5.2	881	1	US-08-333-901-1	Sequence 1, Appl
34	85	5.2	881	1	US-08-456-582-1	Sequence 1, Appl
35	85	5.2	881	3	US-08-898-789-1	Sequence 1, Appl
36	85	5.2	881	3	US-09-039-555B-16	Sequence 16, Appl
37	85	5.2	4544	1	US-08-469-486-52	Sequence 52, Appl
38	85	5.2	4544	2	US-08-469-658-52	Sequence 52, Appl
39	84.5	5.2	2628	2	US-08-570-311-14	Sequence 14, Appl
40	84	5.2	3724	2	US-08-804-227C-10	Sequence 10, Appl
41	84	5.2	3724	2	US-08-804-198-4	Sequence 4, Appl
42	83	5.1	514	3	US-08-600-656-3	Sequence 3, Appl
43	83	5.1	531	3	US-08-688-988-39	Sequence 39, Appl
44	82	5.0	455	2	US-08-870-827-3	Sequence 3, Appl
45	81.5	5.0	557	2	US-08-793-229-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-08-713-298B-2
Sequence 2, Application US/08713298B
Patent No. 5922586
GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dammann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Sch Ielh, Martin
APPLICANT: J rgensen, Per
TITLE OF INVENTION: DNA Constructs and Methods of Producing
TITLE OF INVENTION: Cellulolytic Enzymes
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 59225860 No. 5922586dsk of No. 5922586th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,298B
FILING DATE: 13-SEPT-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3794, 424-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 867-0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-298B-2
Query Match 12.3%; Score 201; DB 2; Length 400;
Best Local Similarity 26.6%; Pred. No. 2e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

```

0Y      8 VSGTLVYANGPEFMRRI--HNGAWY-----KQATTALEGIANGANVRI 54
          | | | | | | | | | | | | | | | | | | | | | | | |
Db      40 ISNGSLVNERGEQVQLKMSHSHGLQWIGQFNYISMETLR-----DMGINFERA 89
          | | | | | | | | | | | | | | | | | | | | | | | |
0Y      55 VL--SDGQWTKDDI--HTVRNLISLAEDELNLVAVPEVHADGYDLSLNRPAVDYIEMR 111
          | | | | | | | | | | | | | | | | | | | | | | | |
Db      90 AMYTSGGYIDDPVSEKYEKKAEVAEALDLDLYIIVDHMLSDNPNTYKEKADFEDEM 148
          | | | | | | | | | | | | | | | | | | | | | | | |
0Y      112 SALLKEPTVLIINEMFEGSMEGDAMADGYK---CAIPLRNAGLHNTLMDVAGWQ 167
          | | | | | | | | | | | | | | | | | | | | | | | |
Db      149 SELVDDYRNVIYELIANENGSS--DYTWQGNQKPLAEVYIPIIRNNDPNNTIIV---GTGT 203
          | | | | | | | | | | | | | | | | | | | | | | | |
0Y      168 FPGSIHDYGREVENADPQRMTFSIHMYEYAGGASOVRTNIDVNLQDIALYIGEGHR 227
          | | | | | | | | | | | | | | | | | | | | | | | |
Db      204 WSDVHF--HAALNQLDLP--NMYIAFHF--YAGTGQNLRDQVDYALDQGAIFVSEWGS 258
          | | | | | | | | | | | | | | | | | | | | | | | |
0Y      228 HTNGD---VDEART--MSYSRGQGVGLAWS 253
          | | | | | | | | | | | | | | | | | | | | | | | |
Db      259 AATGDDGVTFDEDAVYIWFMDERNISMANMS 289
          | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 2
US-08-870-180B-2
; Sequence 2, Application US/08870180B
; Patent No. 5045327

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	12.3%	Score 201	DB 2	Length 400
Best Local Similarity	26.6%	Pred. NO. 2e-11		
Matches 72	Conservative 44	Mismatches 109	Indels 46	Gaps 14

[illegible]

RESULT 3 ;
US-08-814-052-4
; Sequence 4, Application US/08814052
; Patent No. 6015783.

```

1 GENERAL INFORMATION:
2 APPLICANT: von der Osten, Claus
3 APPLICANT: Cherry, Joel R.
4 APPLICANT: Bjornvad, Mads E.
5 APPLICANT: Vind, Jesper
6 APPLICANT: Rasmussen, Michael Dolberg
7 TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
8 TITLE OF INVENTION: OR STAINS FROM CELLULOSE FABRIC
9 NUMBER OF SEQUENCES: 55
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: No. 60157830 No. 6015783dsk of No. 6015783th America, Inc.
12 STREET: 405 Lexington Avenue, Suite 6400
13 CITY: New York
14 STATE: New York
15 COUNTRY: U.S.A.
16 ZIP: 10174-6401
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Diskette
20 COMPUTER: IBM Compatible
21 OPERATING SYSTEM: DOS
22 SOFTWARE: Fastseq for Windows Version 2.0
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/814,052
25 FILING DATE: 06-MAR-1997
26 CLASSIFICATION: 510
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Lambiris, Elias J
29 REGISTRATION NUMBER: 33,728
30 REFERENCE/DOCKET NUMBER: 4684.204-US
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: 212-867-0123
33 TELEFAX: 212-878-9655
34
35 TELEX:
36
37 INFORMATION FOR SEQ ID NO: 4:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 400 amino acids
40 TYPE: amino acid
41 STRANDEDNESS: single
42 TOPOLOGY: linear
43 MOLECULE TYPE: protein
44
45 US-08-814-052-4

```

```

Query Match      12.3%  Score 201,  DB 3,  Length 400,
Best Local Similarity 26.6%  Pred. No. 2e-11,
Matches 72,  Conservative 44,  Mismatches 109,  Indels 46,  Gaps 14,

QY      8  VSGTLLYANGNPFYMKRI-NHGHWY-----KQDATTAGCISANTGTANTYRI 54
      :|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

Db 40 ISNGELVNERGBOVQLKGMSSHG:LOWYGOFWYVESKMYLRD-----DMGINVRA 89

QY 55 VL--SDGGQWTKDDI-HIVRNLI:SLAEDNHLVAVPEVDATGYDSIASLNRAVDYWIEMR 111

Db 90 AMYISSGGYIDDP:SVKEKYEKAEVAIDLDIYI:IDWHLISNDPNITKEEKDFDEK- 148

QY 112 SALIGKEPTVLIININEMFGSWEGDAMADGYK---QAIPRLRNAGLNHTLAVDAAGMQ 167

Db 149 SELIGDYNNVITIEINERNGS--DYTWGNQIKPYAEVPIIRNNDPNITIV---GTGT 203

QY 168 FPOSIDHYGEVEFNADPQRTNFTSIHTEYAGNASQVRTNIDRYLNODLALVIGFGR 227

Db 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLNDQVDYALDQGAALFVSEWGT 258

QY 228 HTNGD---VDPATI-MSYSEORGWGLAMS 253

Db 259 AATGGGVFLDEAQWIDFMDERNLSMANWS 289

RESULT 4

US-08-812-829-4

; Sequence 4, Application US/08812829

; Patent No. 6017751

; GENERAL INFORMATION:

; APPLICANT: von der Osten, Claus

; APPLICANT: Bjornvad, Mads E.

; APPLICANT: Vind, Jesper

; APPLICANT: Rasmussen, Michael Dolberg

; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6017751st of No. 6017751th America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/812,829

; FILING DATE: 06-MAR-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Landlits, Elias J

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 4690, 204-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 400 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-812-829-4

Query Match 12.3%; Score 201; DB 3; Length 400;

Best Local Similarity 26.6%; Pred. No. 2e-11;

Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLLDYDANGNPFVWGI-NHGHAWY-----KDOATTAIGIANTGANTYARI 54

Db 40 ISNGELVNERGBOVQLKGMSSHG:LOWYGOFWYVESKMYLRD-----DMGINVRA 89

QY 55 VL--SDGGQWTKDDI-HIVRNLI:SLAEDNHLVAVPEVDATGYDSIASLNRAVDYWIEMR 111

Db 90 AMYISSGGYIDDP:SVKEKYEKAEVAIDLDIYI:IDWHLISNDPNITKEEKDFDEK- 148

QY 112 SALIGKEPTVLIININEMFGSWEGDAMADGYK---QAIPRLRNAGLNHTLAVDAAGMQ 167

Db 149 SELIGDYNNVITIEINERNGS--DYTWGNQIKPYAEVPIIRNNDPNITIV---GTGT 203

QY 168 FPOSIDHYGEVEFNADPQRTNFTSIHTEYAGNASQVRTNIDRYLNODLALVIGFGR 227

Db 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLNDQVDYALDQGAALFVSEWGT 258

QY 228 HTNGD---VDPATI-MSYSEORGWGLAMS 253

Db 259 AATGGGVFLDEAQWIDFMDERNLSMANWS 289

RESULT 5

US-08-870-180B-13

; Sequence 13, Application US/08870180B

; Patent No. 5945327

; GENERAL INFORMATION:

; APPLICANT: Outtrup, Helle

; APPLICANT: Dammand, Claus

; APPLICANT: Olsen, Arne

; APPLICANT: Bisg, rd-Frantzen, Henrik

; APPLICANT: Sch lein, Martin

; APPLICANT: J rgensen, Per

; APPLICANT: Bjornvad, Mads

; TITLE OF INVENTION: DNA Constructs and Methods of Producing

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5945327th of No. 5945327th America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/870,180B

; FILING DATE: 6-June-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Rozek, Carol E.

; REGISTRATION NUMBER: 36,993

; REFERENCE/DOCKET NUMBER: 3794, 434-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 867 0123

; TELEFAX: 212 867 0298

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 462 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-870-180B-13

Query Match 12.3%; Score 201; DB 2; Length 462;

Best Local Similarity 26.6%; Pred. No. 2.5e-11;

Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLLDYDANGNPFVWGI-NHGHAWY-----KDOATTAIGIANTGANTYARI 54

Db 40 ISNGELVNERGBOVQLKGMSSHG:LOWYGOFWYVESKMYLRD-----DMGINVRA 89

QY 55 VL--SDGGQWTKDDI-HIVRNLI:SLAEDNHLVAVPEVDATGYDSIASLNRAVDYWIEMR 111

DB 90 AAYTSSGGIIDDPSVKEKKEAVEAALDDIYIIDMHLSDNDRIYKEAKDFDEK- 148
QY 112 SALLIGEDTYIINIANEMFGSEGMADGYR---QAIPRLRNAGLNTLMDAAGWQ 167
DB 149 SELYGDPVNIYEIANEPNGS--DVTGNGQIRPYAEVPIIRNDPNIIIV---GTGT 203
QY 168 FPOSIDHGREYFNADPQNTMFSIMYEXAGNASQVNTNDRYLNODLAVIGFGR 227
DB 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGONLNDQVDYDALDGAALFVSEWGS 258
QY 228 HTNGD---VDEATI-MSYSEQGVGWLMS 253
DB 259 AATPDGCVFLDEAQVWIDPMDERNLSWAMWS 289

RESULT 6
US-08-727-548-2
; Sequence 2, Application US/08727548
; Patent No. 5856165
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: No. 5856165el Alkaline Cellulase and Methods of
; TITLE OF INVENTION: Producing Same
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,548
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-727-548-2

Query Match 11.7%; Score 191; DB 2; Length 467;
Best Local Similarity 24.6%; Pred. No. 2,3e-10;
Matches 69; Conservative 49; Mismatches 98; Indels 64; Gaps 14;
QY 8 VSGTTLYPANGNPFYMGRI-NHGAWYKDQATTAEGIANGTGANTVRIVLSDGGQWTRD 66
DB 40 ISNGELVNERGEQVQLKGMSSHGLQWYG-----QFVNYESMKWLDD 81
QY 67 -----IHT-----VKNLISLADNHLVAVPEYHDATGYDSTASLRA 103
DB 82 WGITVFRAMNTSSGGYIIDDPSVKEKKEVEYAAIDIGIYIIDMHLSDNDRIYKEEA 141
QY 104 VDYIEMRSALIGKEDTYIINIANEMFGS---WEGD--AMADGYKQALPRLRNAGLNTL 158
DB 142 KDEFDEM-SELYGDPVNIYEIANEPNGSDVTWQNGIKPYAE---EVLPVIRDNPNIIIV 197
QY 159 AYDAAGWQFPOSIDHGREYFNADPQNTMFSIMYEXAGNASQVNTNDRYLNODL 218
DB 198 IV---GTGTWSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGONLNDQVDYDALDGA 249
QY 219 LVIGFGRHTNGD---VDEATI-MSYSEQGVGWLMS 253
DB 250 IFVSEWGSATATGDCGVFLDEAQVWIDPMDERNLSWAMWS 289

RESULT 7
US-08-924-440-2
; Sequence 2, Application US/08924440
; Patent No. 5871550
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.

APPLICANT: Frits et al.
TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 925 Gate Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/924,440
FILING DATE: August 27, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 846-7555
TELEFAX: (650) 845-6504
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-924-440-2

Query Match 10.5%; Score 171.5; DB 2; Length 430;
Best Local Similarity 24.4%; Pred. No. 1,3e-08;
Matches 77; Conservative 49; Mismatches 122; Indels 67; Gaps 29;
QY 8 VSGTTLYPANGNPFYMGRI-NHGAWYKDQAT-TAIEGIA-NTGANTVRI---VLSGGQ 61
DB 137 VCGIOLDREHNPVQLKGMSSHGLQWYDHCITDSSDLAATWMDADIIIRLSMTIQEGBYE 196
QY 62 WT-----KDDIHTVKNLISLADNHLVAVPEYHDATGYDSTASLNRADVYIEMRSALIGK 117
DB 197 TNPGRFTDRM---QLIDMATARGLYIVDWHILTPDPHYNIDRAKTFEAIHQRRASK 253
QY 118 EDTYIINIANEMFGSEGMADAD--GY-KQALPRLRNAGLNTLMDAAGWQFPOSIDH 174
DB 254 TN-VLYETIAN---PNGVSAISTKSYAEVYIPVIRQDRDPDSVITVTRGWSLIGV-EG 307
QY 175 YGREVENADP--QNTMFSIMYEXAGNASQVNTNDRYLNODLAL---VIGFGRHT 229
DB 308 SGRAEIANAPVNAINIMYAFHF-----ASHDNTLNLARESELFPVYVTEFGIETY 361
QY 230 NGD-----VDEATIMSISYEQGVGWLMSK-----GNGPEMEYIDL 267
DB 362 TGCDANDFQVMDR--YIDLMAERKIGWTKNYSDFRSGAVFQPGTCASGSP----- 411
QY 268 NDWAGNNTLWAGNTI 282
DB 412 --WGSSTLSKSGQWV 424

RESULT 8
US-08-604-913B-11
; Sequence 11, Application US/08604913B
; Patent No. 5712142
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.

APPLICANT: Hammel, Michael E.
APPLICANT: Baker, John O.
TITLE OF INVENTION: METHOD FOR INCREASING
TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: U.S.A.
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II (DOS) text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,913B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,213
FILING DATE: 15-07-1994
ATTORNEY/AGENT INFORMATION:
NAME: Edna M. O'Connor
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: 95-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/384-7573
TELEFAX: 303/384-7499
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FEATURE:
NAME/KEY: E1-CAT
US-08-604-913B-11

Query Match 8.7%; Score 142; DB 1; Length 358;
Best Local Similarity 23.1%; Pred. No. 7.8e-06;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

2 ANSGF-YVSGTILYDANGPFFMARGIN-----HGAMKYDQATTAEGIANAGA 49
1 AGGVTWHTSGREILDANNVPRIAGINWGFETCNTVHG-IMSRY-RSMIDQIKSLGY 58
50 NTVRIYLDGCGMTKDDI---HTVRNLISLADNHLVAVPEYDAGYDSIASLNRADV 106
59 NTRIRLPYS-----DILKPGTMPRSINIFYGMN-----QDLQGLSLGYMKIVAY 103
107 -----WIMRSALT-----GKEDVTIINIANEM 129
104 AGQIGRLIILDRHDPDCSGQSALMTYSSVSEATWISDLOALQRYKGNPTVVGFDLHNE- 162
130 FGSWGDAMADG-----YKQALPRLRNA--GLNHTLMV-----DAAGW----- 165
163 --PHDPACGCGDPSIDWFLAERAGNAVLSVNPILLITVEGVSINGDSYMGGNLQGA 220
166 GQFP-----OSIDHYGREVENADPQRTMFSIHMYEYAGNASQV-RTNIDRVLN 214
221 GQYPVLANVPLNLYVSAHDYATSYV---PQ--TWFSDPFP--NNMPGIWKNNGIYLN 272
215 QDLALY-IGSEFH--RHNGVDYDEATIMSY---SEQRVG---WLAWSKXGK----- 258
273 QAIAPFWGEGFTTLOSTTDQTMATLTVQLRPTAQYGDSPQWTFWSWNPDSGDTGIL 332
259 -PEWEYILSND 269

Db 333 KDDMQTVDTVKD 344

RESULT 9
US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Hammel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 8.7%; Score 142; DB 1; Length 521;
Best Local Similarity 23.1%; Pred. No. 1.3e-05;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

2 ANSGF-YVSGTILYDANGPFFMARGIN-----HGAMKYDQATTAEGIANAGA 49
1 AGGVTWHTSGREILDANNVPRIAGINWGFETCNTVHG-IMSRY-RSMIDQIKSLGY 58
50 NTVRIYLDGCGMTKDDI---HTVRNLISLADNHLVAVPEYDAGYDSIASLNRADV 106
59 NTRIRLPYS-----DILKPGTMPRSINIFYGMN-----QDLQGLSLGYMKIVAY 103
107 -----WIMRSALT-----GKEDVTIINIANEM 129
104 AGQIGRLIILDRHDPDCSGQSALMTYSSVSEATWISDLOALQRYKGNPTVVGFDLHNE- 162
130 FGSWGDAMADG-----YKQALPRLRNA--GLNHTLMV-----DAAGW----- 165
163 --PHDPACGCGDPSIDWFLAERAGNAVLSVNPILLITVEGVSINGDSYMGGNLQGA 220
166 GQFP-----OSIDHYGREVENADPQRTMFSIHMYEYAGNASQV-RTNIDRVLN 214
221 GQYPVLANVPLNLYVSAHDYATSYV---PQ--TWFSDPFP--NNMPGIWKNNGIYLN 272

QY 215 QDLALV-IEEFGH--RHNGDVEATIMSY--SEGRGVC--WLAWSKNG-----258
DB 273 QNAPLWLEEFGLTOSTDQWTKLTVQYLRPTAQYADSEFQWTSWNPNSGDTGIL 332
QY 259 -PEWEYDLSND 269
DB 333 KDDMOTVDTYVKD 344

RESULT 10
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fitch, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59580830 No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 8.3%; Score 135; DB 2; Length 551;
Best Local Similarity 20.9%; Pred. No. 6.8e-05;
Matches 81; Conservative 49; Mismatches 133; Indels 124; Gaps 19;

QY 8 VSGTLLYDANGNPFVARGI-NHGHAWYKDQATTAIEGIANTGANTVRIYISDGGQWTKDD 66
DB 11 VQGNQLVGSGQAVQLVQKMSHGLQWYGN-----FVNRKSLQWNRDN 52

QY 67 --HHYRNLSLADNHNHVAVPEVHDATGDSIASLNR-----VDYWI-----108
DB 53 WGINVRAALYTAEDG-TIDPSVKKNVKEAVASIDLGILYIITDMLISDGNPNYTKAO 111

QY 109 -----EMRSALICEKEDTVIININANWFGSWEGDAMAD--GY-KQATPRLNAGLNIHTIAY 160

DB 112 SKAFQJENATLLGNPNPNIYELANEPNGN---VSNADKSYAEVITAIRADPGCVIY 168
QY 161 DAGMGQFPOSJHDYGREVNADPQRTMFSLHMTYAGNAGNSQVRTNIDRVLDLALV 220
DB 169 GSPPTMSQ-----DIHLADNPVSHSNVYKALHF--YSGTHGQFLDRITVAMNKGALF 220
QY 221 IGEFGHRTNGD-----VDEATIMSYSEGRGVCWLAWS-----WKNGPEWEYL 264
DB 221 VTEMGTSDASGNGGPFYPSQKEMIDFLNARKISWNWMSLADRVETSALMLPGASPTGWT 280
QY 265 DL-----SNDM-----AGN-----NTLAW-GN-----280
DB 281 DAQLSSEGRWYDQIRQATGGSGGNPTAPATNLSATAGNAGVSLTNWAGASITYVK 340
QY 281 --TIYNGPY-----GIRETSRLSTVFT 300
DB 341 RATTSGPYTNVATGVTATSYNTGIL 367

RESULT 11
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC Compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-7

Query Match 6.7%; Score 110; DB 2; Length 531;
Best Local Similarity 19.6%; Pred. No. 0.016;
Matches 54; Conservative 45; Mismatches 118; Indels 58; Gaps 10;

QY 8 VSGTLLYDANGNPFVARGI-NHGHAWYKDQATTAIEGIANTGANTVRIYISDGGQWTKDD 66

DB 44 VGNQLVGGSGAVQLVGGSSHEGLQATGN-----FNKSSLOMARDN 85
QY 67 -IHVRNLISLAEDNHLVAPEVDHATGYDSIASLNR-----VDYWI----- 108
DB 86 WGINVRAMVYSEDDY-ITDPSVKNKVKEAVQASIDLALVYIDHLLSDGNPTYYKAQ 144
QY 109 -----EMRALLGKEDTVIININAMFPGSNEGADGKQALPRLNAGLHTLAVDAA 163
DB 145 SKAFPEMATLYGNTPNVYIEATSPTECVLGGCOSS--EEVTAIRSIDPDGVVIGSP 202
QY 164 GNGQFQSIHDGREFVNADPQNTMFSTHMEYAGGNASQVNTNDRYLNDLAVIGE 223
DB 203 TMSQ-----DHLADNPVSHSNVATLHF--TSGTGGFLDRTITYANKGALFVTE 254
QY 224 FGHRTNGD-----VDEATIMSSEORGVGWGLAWS 253
DB 255 WGTSDASGNGGPPYLPQSKEMIDFLNARKISWVWMS 289

RESULT 12

US-08-745-977-4
; Sequence 4, Application US/08745977
; Patent No. 5770406
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglucanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5770406 No. 5770406disk of No. 5770406th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745, 977
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 350
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4175.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-745-977-4

Query Match 6.1%; Score 100; DB 1; Length 429;
Best Local Similarity 21.9%; Pred. No. 0.11; Mismatches 119; Indels 88; Gaps 18;
Matches 68; Conservative 35; Mismatches 119; Indels 88; Gaps 18;
QY 27 NHGHAMKQATTAIGSIANTGANTVRIYLSDGQWTKDIDHVRN----- 72
DB 106 NHRDWIN--PATVQSVHVGANTIRIP--GYMSYNAIVDTASPPADSNLQLPYLD 159

QY 73 -LISTLAEDNHLVAPEVDHATG---YDSIASLN-----RAVDY-----WIMRSA 113
DB 160 AVYQKADLGIYIILHAPGGQODAFGTGQNPAGFYNTYDGRARKWLSWMTNRH 219
QY 114 LIGKEDTV-IININAMFPGSNEGADW--ADG-----YKQALPRLR-----NAGLN 155
DB 220 TNPATVSGKIEVLENPVSRHGGGRYPAPGDDPSKVQIYYFQALAAVDAEALNVPSN 279
QY 156 HTLVADAGSGQFPOSIDHGREVFVNADPQNTMFSTHMEYAG-----GNASQVR--- 206
DB 280 KRLHV-----QMSKMSGDPDSNAAYKNDPMVEFDHNTYIGFALQNTGDDYSLMHS 333
QY 207 -INIDRYLNDLALVYGEGRHTNG-DVDEATIMS-----YSEORGVGWGLAWSK 255
DB 334 CIDSRYVSGGDFP-ITGEMS--MTSGADWHDGNEFTKFTTAQOOLYESPGMDGMIYTWK 390
QY 256 G--NGPEWEY 263
DB 391 TELNDPRWTY 400

RESULT 13

US-09-040-699A-4
; Sequence 4, Application US/09040699A
; Patent No. 6022723
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglucanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6022723 No. 6022723disk of No. 6022723th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040, 699A
; FILING DATE: 18-March-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valera A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4175.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-040-699A-4

Query Match 6.1%; Score 100; DB 3; Length 429;
Best Local Similarity 21.9%; Pred. No. 0.11; Mismatches 119; Indels 88; Gaps 18;
Matches 68; Conservative 35; Mismatches 119; Indels 88; Gaps 18;
QY 27 NHGHAMKQATTAIGSIANTGANTVRIYLSDGQWTKDIDHVRN----- 72

```

DB 106 NMYRMIN---PATVSHADVLANTRIRPI-----GYMSNAYVDTASEPFDQNLQPLYL 159
QY 73 -LISTAEDNHLVAVPEVDATG---YDSTASLN-----RAVDY-----WLEKMSA 113
DB 160 AVQKADIGIYIIDLHGAPEGQOQDAFTGQNPAGFYNTYDGRAEKMLSWMTNFIH 219
QY 114 LGKEDEV-IINANEMSGWEGDAM--ADG-----YKQALPRLR-----NAGLN 155
DB 220 TNPASTVMIEVLEPEVSRHGGCHYPAFGQDPSSVQIYFGALKAVDAAALNVP SN 279
QY 156 HTLMDVAMGQFOSIHIDYGREVNADPQRTMFSIMHYEAG-----GNASQVR--- 206
DB 280 KKLHV-----QFMSKMSGSDPRSNAAYKNDPMVGFDDHNYIGFALQNTGQYSILMSA 333
QY 207 -TIDRYLMDLALVYGECHRTNG-DVDEATIMS-----YSEBGVGLAWSK 255
DB 334 CTDSTRVSGQDFA-ITGEMS--MTSGADWHDGNFTKFTTAQOQLYESPGMDGWLITWK 390
QY 256 G--NGPEWEY 263
DB 391 TELNDPRWTY 400

```

RESULT 14

```

US-09-066-075-2
; Sequence 2, Application US/09066075
; Patent No. 5925749

```

GENERAL INFORMATION:

```

APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
NUMBER OF SEQUENCES: 4

```

CORRESPONDENCE ADDRESSES:

```

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,075
FILING DATE:

```

CLASSIFICATION:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615
FILING DATE: August 23, 1995

```

ATTORNEY/AGENT INFORMATION:

```

NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1744

```

INFORMATION FOR SEQ ID NO: 2:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 317 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-066-075-2

```

```

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

```

```

QY 19 NPFVARGINRGA-----W---YKDAQTLAEGANTGANTVRIYVSDSGQWTK--- 64
DB 9 NKILGRINIGNMLEAPNEGDWGVYIDE--PFDIKRKGFSHVRIPI---RWSITHAV 61
QY 65 -----DDI--HTVNLISLAEDNH---LVAVEVDATGSDYSIASINRA 103
DB 62 AEPPIKMDPEFKVADVINGALKRGALVALNHHYELANDNDEBEK-----ERF 111
QY 104 VDIWENRSALIGKEDVIINIANEMFSGWEGDAMADGYQALPRLNAGLNHTIAMDAA 163
DB 112 LALKQJADRYKDYETLFEILNEPHGNLPEKWNELLEALKVIRSIDKKHIIIGTA 171
QY 164 GNGQFPQ---SIDYGREVNADPQRTMFSIMY---EYAGSNASQVNTIDRYLND 216
DB 172 EMGGSIALEKLSVRKW-----EKNSYVTHIYNPFETHOGALWEGS--EKWIGRK 221
QY 217 LA-----LVIGFCHRTNGDVDE---ATINSYSGRG 246
DB 222 WGSPPDQKHLIEFNFIEMSKKKRPIYIGFEG-AVRKADLESRIKWTSPVAREMEKR- 279
QY 247 VGWLAWSKNGPEWEY 263
DB 280 ----RMSW---AYWEF 288

```

RESULT 15

```

US-08-518-615A-2
; Sequence 2, Application US/08518615A
; Patent No. 5962258

```

GENERAL INFORMATION:

```

APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
NUMBER OF SEQUENCES: 4

```

CORRESPONDENCE ADDRESSES:

```

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615A
FILING DATE: August 23, 1995

```

CLASSIFICATION:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

```

ATTORNEY/AGENT INFORMATION:

```

NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744

```

INFORMATION FOR SEQ ID NO: 2:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 317 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-518-615A-2

```

```

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

```

```

QY 19 NPEVARGINHGHA-----W---YKDOATTAIEGIANTGANTVRIYLSGGGOWTK--- 64
| : ||| | : |
Db 9 NKILRGINIGNALZAPNEGDMGVYKDE--FDIIEAGFSHVRIPI---RWSTHAY 61
| : ||| | : |
QY 65 -----DDI--HTVRNLISLAEDNH---LVAVPEVHADAGYDSIASLNRA 103
| : ||| | : |
Db 62 AEPFYKINDRPFKRYDEVINGALTKRGIAVALINHHYBELMNDPEHK-----ERF 111
| : ||| | : |
QY 104 VDYWIEMRSALIGKEDTVIINIANEMFGSWEGDANADGYKQAIPLRNAGLNHTLMVDA 163
| : ||| | : |
Db 112 LALMFOIADRYKDYPETLFEFELNPHGMLPEKWNELLEALKYIRSIDKHTIIGTA 171
| : ||| | : |
QY 164 GMSGPQ-----SIHGYREVFNADQORNTMSIHMV---EYAGNASQVRTNIDRYLND 216
| : ||| | : |
Db 172 EMGGISALEKLSVPRW-----EKNSVTIHTYINPFEETHGAEWVEGS-EKMLGRK 221
| : ||| | : |
QY 217 LA-----LVIGFGHRTNGDYDE-----ATIMSYSEORG 246
| : ||| | : |
Db 222 WGSPPDOHLEEFNFIEEMSKNKRPIYIGEPG-AYKRADESRKWTSPVREMEKR- 279
| : ||| | : |
QY 247 VGWLAWSWKGNGPEWEY 263
| : ||| | : |
Db 280 ----RMSW----AYWEF 288
| : ||| | : |

```

Search completed: December 19, 2000, 17:02:47
 Job time: 1279 sec

This Page Blank (uspio)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:35:51 ; Search time 132.02 Seconds
(without alignments)
72.598 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 NANSGFYVSGTTLTYDANGNP.....TIVNGPYGLRETSRLSTVF 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692	42.5	383	1 MANA_STRLI	P51529 streptomyc
2	601	36.9	1331	1 MANB_CALSA	P25533 caldoceim
3	202	12.4	409	1 GUN2_BACSA	P06565 bacillus sp
4	201	12.3	400	1 GUN5_BACAG	P06566 bacillus ag
5	200	12.3	488	1 GUN1_BACSA	P06566 bacillus sp
6	184.5	11.3	505	1 GUNV_ERWCA	Q47096 erwilia car
7	180.5	11.1	444	1 GUNV_ERWCA	Q59394 erwilia car
8	180	11.0	448	1 GUN_CLOAB	P15704 clostridium
9	177.5	10.9	466	1 GUN5_THIEPU	Q01786 thermomonos
10	174.5	10.7	504	1 GUNV_ERWCA	Q59395 erwilia car
11	173.5	10.6	499	1 GUN1_BACSU	P07983 bacillus su
12	162.5	10.0	499	1 GUN2_BACSU	P10475 bacillus su
13	161	9.9	459	1 GUN1_STRLI	P27035 streptomyc
14	152	9.3	800	1 GUN1_BACSA	P06566 bacillus sp
15	151.5	9.3	499	1 GUN3_BACSU	P23549 bacillus su
16	151.5	9.3	825	1 GUN3_BACSA	P19570 bacillus sp
17	149.5	9.2	941	1 GUN_BACSA	P19424 bacillus sp
18	147.5	8.7	429	1 GUN1_BACSA	P22541 butyrylhydrol
19	142	8.7	562	1 GUN1_ACICE	P54583 acidotherm
20	141	8.7	312	1 GUN4_RUMAL	Q07940 ruminococcu
21	131	8.0	814	1 GUN2_CLOTH	P10477 clostridium
22	130	8.0	426	1 GUN_BACSA	P17974 burkholderi
23	122.5	7.5	1039	1 GUNB_CALSA	P10474 c endoglucan
24	122	7.5	438	1 EXG_CANAL	P29717 candida alb
25	117	7.2	426	1 GUN2_ERWCH	P07103 erwilia chr
26	115.5	7.1	747	1 GUN2_CELPFI	P50400 cellulomona
27	113.5	7.0	517	1 GUN4_CLOLO	P54937 clostridium
28	111	6.8	475	1 GUN4_CLOCE	P17901 clostridium
29	109.5	6.7	440	1 GUNB_CLOCL	P23660 ruminococcu
30	105	6.4	364	1 GUN4_RUMAL	P06566 bacillus su
31	104	6.4	566	1 GUN2_CLOTH	P27035 pseudomonas
32	100.5	6.2	748	1 GUNC_PSEFL	P23776 pseudomonas
33	99.5	6.1	448	1 EXG1_YEAST	P23776 pseudomonas

34	98	6.0	584	1 GUN2_CLOCE	P25472 clostridium
35	97.5	6.0	375	1 GUN_FOBSBP	P23044 robilliarca
36	97.5	6.0	647	1 CN16_ECOLI	P08331 escherichia
37	97	6.0	493	1 GUN4_XANCP	P19487 xanthomonas
38	95	5.8	515	1 GUN2_CLOCL	P28623 clostridium
39	94.5	5.8	645	1 LIPI_PHOLO	P40601 photorhabdu
40	94.5	5.8	856	1 AAP1_YEAST	P37898 saccharomyc
41	94	5.8	562	1 EXG2_YEAST	P52911 saccharomyc
42	94	5.8	890	1 BCN5_CLOPE	P08696 clostridium
43	93.5	5.7	406	1 GUN1_RUMAL	P16216 ruminococcu
44	93.5	5.7	510	1 YDYL_SCHMO	O13692 schizosacch
45	93.5	5.7	566	1 GUNB_PABLA	P23550 penibacilli

ALIGNMENTS

```

RESULT 1
ID MANA_STRLI STANDARD; PRT; 383 AA.
AC P51529;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MANNAN ENDO-1,4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-
DE MANNANASE) (1,4-BETA-D-MANNAN MANNANOHYDROLASE).
GN MANA.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 36-42.
RC STRAIN-66 / 1326;
RX MEDLINE; 93207541.
RA Arcand N., Riuepfeil D., Paradis F.W., Morosol R., Shareck F.;
RT "Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence
RT of the manA gene and characterization of the enzyme.";
RL Biochem. J. 290:857-863(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN-66 / 1326;
RA Shareck F.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES
CC CELSIUS AND A PH OF 6.8.
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-
CC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
CC GALACTOGLUCOMANNANS.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M92297; AAA26710.2; -.
CC INTERPRO; IPR001547; -.
CC PFM; PF00150; cellulase; 1.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC HYDROLASE; GLYCOSIDASE; Signal.
CC SIGNAL
CC FT 1 35
CC CHAIN 36 383 MANNAN ENDO-1,4-BETA-MANNOSIDASE.
CC FT 336 340 POLY-GLY.
CC DOMAIN 383 AA; 39681 MW; SDB4B407C64E94C3 CRC64;
CC SEQUENCE

```

Query Match 42.5%; Score 692; DB 1; Length 383;
Best Local Similarity 46.7%; Pred. No. 2,3e+46;

Matches 141: Conservative 51: Mismatches 104: Indels 6: Gaps 5:

QY 2 ANSGTYVSTIYDANGNPTVRKINHGHWTKDQATTAIEGANTGANTVIVISDGGQ 61
 DB 36 AAGGIVHSNGRVEENGSGSFVRGVNHAFTWYPPDR-TGSDIADIAKGAQTVAVVLSGGR 94
 QY 62 WTKDDIHVFNRLISLAEDNHLVAPEVHDATGY--DSIASLNRADVNIEMKMSAL-ICE 118
 DB 95 WTKTASAEVSLIGOCKANKVYCEVHDHTYIGDQATSLQAGDYVWYGAASAMRAOE 154
 QY 119 DTVIINIANEWGSEGDAMADYKQALPRLNAGLNLHTLVDAAGWQ-EPQSIHDYGR 177
 DB 155 DVVVNVIGNEPGNTNYAAMTATKSAIGKLKAGLHMLAVDAPWQDMGSGTRSNAA 214
 QY 178 EYFNADPQRTMFSTHMEYAGNASQVRTINDRYLNDLALVIEFGHRTNGVDENAT 237
 DB 215 SYFASDPQRTNYFSIHMVG-VYDTLAEVNDYINAAVGNGLPVVEGFGHSDGDDPDEDA 273
 QY 238 IMSYEQRGVGLWMSKNGSGPEWEXYLDLSNDWAGNNLTAVNGTYNGPYGLRSTRSLT 297
 DB 274 IMATQSLGCVGLGMSGNGGVEYLDWYNGFDPNSLITSMGNRLIYSGNIAATSRAT 333
 QY 298 VF 299
 DB 334 VY 335

RESULT 2
 MANB_CALSA STANDARD; PRT; 1331 AA.

AC P22533;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BETA-MANNANASE/ENDOGALACTANASE A PRECURSOR (INCLUDES: MANNAN ENDO-1,4-
 DE BETA-MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-
 DE MANNANASE); ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE)].
 GN MANA.
 OS Caldocellum saccharolyticum (Caldicellulosigrupor saccharolyticus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermotomobacter group; Caldicellulosigrupor.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93119139.
 RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
 RT multidomain enzyme.";
 RL Appl. Environ. Microbiol. 58:3864-3867 (1992).
 RN [2]
 RP SEQUENCE OF 1-346 FROM N.A.
 RX MEDLINE; 91247819.
 RA Luecht E., Jesmet N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in *Escherichia coli* of a
 RT gene coding for a beta-mannanase from the extremely thermophilic
 RT bacterium 'Caldocellum saccharolyticum'.";
 RL Appl. Environ. Microbiol. 57:694-700 (1991).
 CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT
 CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
 CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC
 CC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
 CC GALACTOGLUCOMANNANS.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
 CC CELSIUS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on the
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; L01257; AAA71887.1; -;
 DR EMBL; M35063; AAA72861.1; -;
 DR PIR; B43745; B43745.
 DR PIR; A48954; A48954.
 DR HSSP; Q06851; INBC.
 DR INTERPRO; IPR001547; -;
 DR INTERPRO; IPR001956; -;
 DR PRAM; PF00942; CBD_3; 2.
 DR PFAM; PF00150; cellulase.1.
 DR PROSITE; PS00659; GLYCOSYL-HYDROL_5; 1.
 DR Hydroxylase; Glycosidase; Cellulose degradation; signal;
 KW Multifunctional enzyme.
 FT SIGNAL 1; 41
 FT CHAIN 42; 1331
 FT DOMAIN 42; 325
 FT DOMAIN 326; 361
 FT DOMAIN 362; 518
 FT DOMAIN 519; 564
 FT DOMAIN 565; 720
 FT DOMAIN 721; 780
 FT DOMAIN 781; 1331
 FT ACT_SITE 162; 162
 FT ACT_SITE 257; 257
 FT ACT_SITE 338; 338
 FT CONFLICT 340; 346
 FT SEQUENCE 1331; AA; 146892 MM; FFBGA51B8BDF0E0 CRC64;

Query Match 36.9%; Score 601; DB 1; Length 1331;
 Best Local Similarity 44.5%; Pred. No. 1.4e-38;
 Matches 125; Conservative 42; Mismatches 108; Indels 6; Gaps 4;

QY 25 GINHGHWYKQATTAIEGANTGANTVIVISDGGQWTDHHTVNLISLAEDNHLVA 84
 DB 55 GTNNAHGWYRDLDTALRGIRSGMNSVRYVLSNGYMWIRIPASEVANIISLSGFKR 114
 QY 85 -VPEVHDATGY--DSIASLNRADVNIEMKMSALIGKEDYIINIANEWGSEGDAMAD 140
 DB 115 IILEVHTTQYGEAGACSLAQAVRYKEIKSVLDGHEDEYIINIGEPYGNNTQWVNV 174
 QY 141 GYKQALPRLNAGLNLHTLVDAAGWQ-EPQSIHDYGRVFNADPQRTMFSTHMEYAG 199
 DB 175 DTKNAIKALRAGFKRTIMVADAPWQDMGTMRDNASINEADPLNLVFSIHMVG-VY 233
 QY 200 GMSQVPTINDRYLNDLALVIEFGHRTNGVDENATIMSSEQRGVGLWMSKNGNP 259
 DB 234 NTASKVEEYIKSFVQDKLPVIEFGHQHTDGDDEEATVYAKQYKIGLFSWCGNSS 293
 QY 260 EWEYLDLSNDWAGNNLTAVNGTYNGPYGLRSTRSLTVP 300
 DB 294 YVGILDMVNNWDRNPNTPMGQWKINAIGSSPTPISTVY 334

RESULT 3
 GUN2_BACS4 STANDARD; PRT; 409 AA.

ID GUN2_BACS4
 AC P06565;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE).
 GN CELB.
 OS Bacillus sp. (strain N-4).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE: 87056924.
RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
RT "Nucleotide sequences of two cellulase genes from alkalophilic
RL Bacillus sp. strain N-4 and their strong homology."
RL J. Bacteriol. 168:479-485(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M14729; AAA22299.1; -.
DR PIR: B25156; B25156.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KW Cellulose degradation; Hydrolase; Glycosidase.
FT ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 409 AA; 43690 MW; 988860686B4DA3F CRC64;

Query Match 12.4%; Score 202; DB 1; Length 409;
Best Local Similarity 25.0%; Pred. No. 1.6e-08;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

OY 8 VSGTLLDANGNPFYMRGI-NHGMVYKQATTAEGTANTGANTVRIYLSGGQWTKDD 66
DB 40 ISNGELVNRGEPOVLKGMSSHGLOWYGFVNYESMKWLRD 81
OY 67 -----IHT-----VNLISLADNHLVAVPEVDATGYSTISLARA 103
DB 82 WGITVFRAMVITSSGGYIEDPSKREKVEVAIDLGIYVIDHIIISDNDPNITKREA 141
OY 104 VDYWIEKMSALIGKEDVTYIINANEMFGS---WEGD--AMADGYKQATPRLNAGLNTL 158
DB 142 KDFPEDEM-SELXGDPNPTIETIANEPNSDVTWNOIKPYAE---EVIPVIANNDPNNTI 197
OY 159 MVDAAGWQGFPOSHIDYREVFNADPQRTNMFISIMTEYAGGNSAQVRTNIDRVLNODLA 218
DB 198 IV---GTGTWSQDVH-HAADNQLTDP--NVMYAFHF--YAGTHGQNLNDQVDYALDQGA 249
OY 219 LVIGEFGRHRTNGD---VDENIT-MSYSEORGVGWLAWS 253
DB 250 IFVSEMGTSSEATGDGCVFLDEAOWIDFMDERNLSWANWS 289

RESULT 4
GUNS_BACAG STANDARD; PRT; 400 AA.
AC 085465;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENDOGUCANASE 5A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (ALKALINE
DE CELLULOSE).
GN CEL5A.
OS Bacillus agaradhaerens (Bacillus agaradhaerens).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 8721;
RA Bjornvad M.E.;
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
RN [2]

```

```

RP X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329.
RC STRAIN-AC13 / NCIMB 40482;
RX MEDLINE: 98153671.
RA Davies G.J., Dauter M., Brzozowski A.M., Bjornvad M.E.,
RA Andersen K.V., Schuelein M.;
RT "Structure of the Bacillus agaradhaerens family 5 endoglucanase at 1.6-
RT A and its cellobiose complex at 2.0-A resolution.";
RL Biochemistry 37:1926-1932(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329.
RC STRAIN-AC13 / NCIMB 40482;
RX MEDLINE: 98384136.
RA Davies G.J., Mackenzie L.F., Varrot A., Dauter M., Brzozowski A.M.,
RA Schuelein M., Withers S.G.;
RT "Snapshots along an enzymatic reaction coordinate: analysis of a
RT retaining beta-glycoside hydrolase.";
RL Biochemistry 37:11707-11713(1998).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -----
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF067428; AAC19169.1; -.
DR PDB: 1A3H; 16-MAR-99.
DR PDB: 2A3H; 16-MAR-99.
DR PDB: 3A3H; 16-MAR-99.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACT_SITE 165 165 PROTON DONOR.
FT ACT_SITE 254 254 NUCLEOPHILE.
SQ SEQUENCE 400 AA; 44702 MW; 3F9C66FB9B36FF CRC64;

Query Match 12.3%; Score 201; DB 1; Length 400;
Best Local Similarity 26.6%; Pred. No. 1.8e-08;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

OY 8 VSGTLLDANGNPFYMRGI-NHGMVY-----KQATTAEGTANTGANTVRI 54
DB 40 ISNGELVNRGEPOVLKGMSSHGLOWYGFVNYESMKWLRD-----DMGINVERA 89
OY 55 VL--SDGGQWTKDDI-HVIRNLIISLADNHLVAVPEVDATGYSDSIASLNRAVDWIMER 111
DB 90 AMYTSSGGYIDDPYKREKVEVAIDLGIYVIDHIIISDNDPNITKREKDFDEM- 148
OY 112 SALIGKEDVTYIINANEMFGSWEADADGYK---QAIPRLNAGLNTLAVDAAGWGO 167
DB 149 SELXGDPNPTIETIANEPNS--DVTWNOIKPYAEVPIIRNNDPNNTIIV--GTGT 203
OY 168 FPQSHIDYREVFNADPQRTNMFISIMTEYAGGNSAQVRTNIDRVLNODLALVIGEFGR 227
DB 204 WSQDVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLNDQVDYALDQGAIFVSEMGTS 258
OY 228 HTNGD---VDENIT-MSYSEORGVGWLAWS 253
DB 259 AATGSGVFLDEAOWIDFMDERNLSWANWS 289

RESULT 5
GUNI_BACS4 STANDARD; PRT; 488 AA.
AC P06566;

```

```

DN 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
GN CELA.
OS Bacillus sp. (strain N-4).
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87056924.
RA Fukumori F., Sasahara N., Kudo T., Horikoshi K.;
RT "Nucleotide sequences of two cellulase genes from alkalophilic
RT Bacillus sp. strain N-4 and their strong homology."
RL J. Bacteriol. 168:479-485(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: M14781; AAA2301.1;
DR PIR: A25156; A25156.
DR INTERPRO: IPR001547;
DR PRFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR KMW CELLULOSE degradation; Hydrolase; Glycosidase.
FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 488 AA; 54264 MW; 97248E453D28D3B5 CRC64;

```

Query Match 12.3%; Score 200; DB 1; Length 488;
 Best Local Similarity 24.6%; Pred. No. 2,8e-08;
 Matches 70; Conservative 49; Mismatches 103; Indels 62; Gaps 14;

```

OY 3 NSGFYSGTGLDANGNPPVNGI-NHGHWYKQDATTAEIGANTGATVRLSDGG 61
DB 33 NGQLSTQNGQLVNEHGDVQLKGMSSHGQWYG-----PQVNTDSTK 74
OY 62 WTKDD-----IHT-----YRNLISLAEDNHLVAVPEVHATGYDSTA 98
DB 75 WTRDDMGIVFRAMATSSGGYEDSVKRYEAVENALDGIYIIDMHLSDNDPNT 134
OY 99 SLNRADVWIEMRSALIGKEDYIINANWFG---SWGDAMADGY-KOALPRLNAGL 154
DB 135 YKEAEKEFDEEM-SALYGYPNVYIENEPNGHNVRW--DSHIKPYAEVYIVIRANDP 191
OY 155 NHTLWADAGMGQFQSIDYGRVFNADPQRTMFSIMHYEAGGASQVRRINDRYLN 214
DB 192 NNTVIVIGTAWTS---ODVHEADNDQLD-DP--NVMYAFHF--YAGTHGOQLRQVYALIS 243
OY 215 QDLATIGEGHRTNGD---VDEATI-MSYSEQRGVGLWANS 253
DB 244 RGAATFVSEMGTSATGSDGVFLDEAQWTDVDFEDRLNLSMANS 287

```

RESULT 6
 GUNV_ERMCA STANDARD; PRT; 505 AA.

```

AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ENDOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)
DE [CELLULOSE V].

```

```

GN CELV.
OS Erythra carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SCRI193.
RX MEDLINE: 94067016.
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (celV) of Erythra
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains."
RL Mol. Gen. Genet. 241:341-350(1993).
CC -1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
CC OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
CC CELSIUS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: X76000; CAA3592.1;
DR HSSP: 006851; INBC.
DR INTERPRO: IPR001547;
DR INTERPRO: IPR001956;
DR PRFAM: PF00942; CBD_3; 1.
DR PRFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR KMW CELLULOSE degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 505 ENDOGLUCANASE V.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

```

Query Match 11.3%; Score 184.5; DB 1; Length 505;
 Best Local Similarity 23.1%; Pred. No. 4.6e-07;
 Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

```

OY 13 LYDANGNPFVNGI-NHGHWYKQDATTAEIGANTGATVRLSDGGQWTKDD----- 65
DB 48 LVDEQGRVQLRGISSHGQWFGD-----YVKKDSKMTLRDDMGGINV 89
OY 67 -----IHYRNLISLAEDNHLVAVPEVHATGYDSTAENRAVDYWI 108
DB 90 FVAVMTAADGYISNPISLANKEAVAAQSLGYITIIDMHLSDNDPNTYRAQAKTFEA 149
OY 109 EMRSALIGKEDYIINANWFG---SWGD--AMADGYKOALPRLNAGLNTLWADAG 164
DB 150 EM-AGIGSSPNVYIENEPNGVITWNGQIRPYA---LEVYDITRSNDPDLIIV---G 202
OY 165 WGFQPSIDHYGRVFNADPQRTMFSIMHYEAGGASQVRRINDRYLNQDLALVIGEF 224
DB 203 TGTWSDIHD---AADNQPDPTMYALHF--YAGTHGOFLDRIDYQASRGAAIFVSEW 257
OY 225 GARTHNGD---VDEATIMSYSEQRGVGLWANSKNGKNGPEHYL-----DLSNDAGNLT 275
DB 258 GTSDASNGNGPPLPESQWTIDFLNNGVSWVMSLTDNSEASALAPGASKSGGWTQNTL 317
OY 276 TAWGNTI 282

```


DB 159 AEE-YGKSYNIYEICNEPNC---GTMWANDIKRYANYIIIPALIPDNNIIIVGTSTWS 214
 QY 167 QPPOSIMHYGAEVFNADQR--NTMESIHMEYAGNASOVRTNIDRYLNDLALVGEF 224
 DB 215 QDVADIAD-----NPLRSNIMTYCHF--YAGTHQSILRDKINAMSGIALFYEW 264
 QY 225 GHRRTNCD-----VDEA-TIMSYSOERGVMGLAWSKNGPMEYXL---DISNDWAGNNL 275
 DB 265 GTSDBASGNGPFLDBSQKWDVPMASKNISWTNNWALCDKSEASALKSGSSTTGGWTDSDL 324
 QY 276 TAWGNTIVNGPYGLRETSRLS 296
 DB 325 TTSGLEFVKRSIGSGSNTSOTS 345

RESULT 9
 GUNS_THERFU STANDARD; PRT; 466 AA.
 ID GUNS_THERFU 001786;
 AC 01-JUL-1993 (Rel. 26, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4) (CELLULOSE E-5) (CELLULOSE E5).
 GN CELE.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangiales; Nocardiopsaceae; Thermobifida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YX;
 RX MEDLINE; 91258320.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RT "DNA sequences of three beta-1,4-endoglucanase genes from
 Thermomonospora fusca".
 RT J. Bacteriol. 173:3397-3407(1991).
 RN [2]
 RP REVISIONS.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RT Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 137-142 AND 157-166.
 RC STRAIN-YX;
 RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
 RT "Activity studies of eight purified cellulases: specificity,
 synergism, and binding domain effects".
 RT Biotechnol. Bioeng. 42:1002-1013(1993).
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -1- PATHWAY: CELLULOSE DEGRADATION
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: L01577; AAC09379.1; -
 DR PIR: C42360; C42360.
 DR HSSP: P07986; 1EXH.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001919; -
 DR PRAM: PR000553; CSD.2; 1.
 DR PRAM: PR00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosylase; signal.
 FT SIGNAL 1 36
 FT CHAIN 37 466 ENDOGLUCANASE E-5.

FT ACT SITE 299 299 PROTON DONOR (BY SIMILARITY).
 FT ACT SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 466 AA; 49800 KM; 1CF0ADF2DEF82E CRC64;
 Query Match 10.9%; Score 177.5; DB 1; Length 466;
 Best Local Similarity 24.8%; Pred. No. 1,4e-06;
 Matches 78; Conservative 49; Mismatches 121; Indels 67; Gaps 18;
 QY 8 VSGTLLDANGNPFVAMGIN-HGHAWYKQAT-TALEGIA-NTGANTVRL---VLSGGQ 61
 DB 173 VCGTQLCDENGNPVLQRMSTHGIOWFDHCLTDSLDALVDKADILRLISMVIOEGYE 232
 QY 62 WT-----KDDIHTVANKLSLADNHLVAVPEVHATGDSJASINRAVDYIEKRSALIG 117
 DB 233 TNPRTGFTDRM---QLIDMNTARGLYIVDMHILTPEDPHYNDPRATFTFAELTAQRASK 289
 QY 118 EDTVILINANEMFEGSWEGDAD--GY-KOALPRLNAGNHTLMDVDAAGWGFPOSIND 174
 DB 290 TN-VLEYLANE---PQGVSWASIKSYAEVYIPYIRQDPDSVYIVGRGWSLGVN-EG 343
 QY 175 YGREVFNADP--QRTMFSIHMEYAGNASOVRTNIDRYLNDLAL---VIGFGRHT 229
 DB 344 SGPAEIAANPNAASINIMYAEFFY-----AASHRDNYLNLREASELFPVFTVEFTET 397
 QY 230 NGD-----VDEATIMSYSOERGVMGLAWSK-----GNPMEYIDL 267
 DB 398 TGDANDFQMDR--YIDLMAERKIGMTKNYSDFRSAGVFORGTCASSGP----- 447
 QY 268 NDWAGNNLTANGNTI 282
 DB 448 --WSSSLKASGQWV 460

RESULT 10
 GUNW_ERMCA STANDARD; PRT; 504 AA.
 ID GUNW_ERMCA 059395;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ENDOGLUCANASE VI PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE VI) (CELLULOSE VI).
 GN CELV1.
 OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SCC3193;
 RX MEDLINE; 95231512.
 RA Mae A., Heikkinheimo R., Palva E.T.;
 RT "Structure and regulation of the Erwinia carotovora subspecies
 carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 phytopathogenicity".
 RT Mol. Gen. Genet. 247:17-26(1995).
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X79241; GAA55823.1; -
 DR HSSP: O06851; 1NRC.
 DR INTERPRO: IPR001547; -

```

DR INTERPRO: IPR001956;
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR CELLULOSE degradation; Hydrolase; Signal.
KW SIGNAL 1 31
FT CHAIN 32 504
FT DOMAIN 32 334
FT DOMAIN 335 352
FT DOMAIN 353 504
FT ACT_SITE 168 168
FT ACT_SITE 256 256
FT ACT_SITE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;
SQ SEQUENCE

Query Match
Best Local Similarity 10.7%; Score 174.5; DB 1; Length 504;
Matches 73; Conservative 43; Mismatches 123; Indels 69; Gaps 14;

Db 13 LYDANGNPFVWARGI-NHGHAMVKDQATTAEGIANGTANTYRIYLSGGQWTKD--IHT 69
48 LVDSGRKRVQLRGSSNGLOWGD-----YNNKDSMKWLRDWDGIVN 89
QY 70 VRLNLSIAEDNHLVAVEV-----HDATGYDSIASIRAVDYW 107
Db 90 FRVAMYTAE-NGYIANSLANKVEAAAGLGVIYIIIDHILSDNDPNTYKQAKIFF 148
108 IEMRSALIGKEDYVIINIANEWFGS--WEGD--AMADGYKQAIPLRNAGINHTLVADA 163
QY 149 AEM-AGLYGSPNYIETIANEPNGSVTWNGQIRPYA--LEVTDTIRSKDNDLIIYGS 204
164 GWGQPSIDHYGREVNADPQRTMFSIHMYEYAGNAGSVRNIDRYLNODLALVIGE 223
QY 205 TWS---ODIHD---ADNDQLEDPPTLYALHF--YAGTHGQFLRDRIYDAGSGAIVPSE 256
Db 224 FGHRTNGD-----VDEATIMSYSEORGCVGLWAMSKNGPEMYL---DLSNDMAGNN 274
257 WGTSDASGNGPFLPESQWTWDFLNNRGISIVNKSLSDKSTETSAALVAGASKSGWTEQN 316
QY 275 LTWAGNTI 282
Db 317 LSTSGKEV 324

RESULT 11
GUN2_BACSU STANDARD; PRT; 499 AA.
ID GUN2_BACSU
AC P07983;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN BGIC OR GLD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DIG;
RC MEDLINE: 87194581.
RA Robson L.M.; Chambliss G.H.;
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
RL J. Bacteriol. 169:2017-2025(1987).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: M16185; AAA22496.1; ALT_INIT.
DR PIR: A26874; A26874.
DR HSSP: Q06851; INBC.
DR INTERPRO: IPR001547;
DR INTERPRO: IPR001956;
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR CELLULOSE degradation; Hydrolase; Glycosylase; Signal.
KW SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169
FT ACT_SITE 257 257
FT DOMAIN 350 499
SQ SEQUENCE 499 AA; 55187 MW; 339D04EE95A63EE1 CRC64;

Query Match
Best Local Similarity 10.6%; Score 173.5; DB 1; Length 499;
Matches 69; Conservative 68; Mismatches 122; Indels 73; Gaps 15;

QY 3 NSGFYVSGTIIYDANGPFVWARGI-NHGHAMVKDQATTAEGIANGTANTYRIYLSGGQ 61
39 NGQLSIRGTQLVNDGKAVOLKGISSHGLWYG-----FNNKSKLK 80
QY 62 WTKD-----IHT-----VRLNLSIAEDNHLVAVEVDATGYDIA 98
Db 81 WLRDWDGIVFRAMYTADGVIIDNPVYKAVKAVAKLGIYIYIIDHILSDNDPNTYKQAKIFF 140
99 SLNRAVDYWIEMRSALIGKEDYVIINIANEWFG--WEGD--AMADGYKQAIPLRNAGL 154
Db 141 NKEAKKEFFKEM--SLVGNPNVYIETIANEPNGDVNKRDKIPYAE--EIVSVIRKNDP 196
QY 155 NHTLVADAAGNGQFPOSIDHYGREVNADPQ--RTMFSIHMYEYAGNAGSVRNIDR 211
197 DNIITV--GTGTSQDVND-----AADQLEKPDANVYALHF--YAGTHGQSLRDANY 245
QY 212 VLNODLALVIGEFGHRTNGD---VDEA--TIMSYSEORGCVGLWAMSKNGPEMYL-- 264
246 ALSKGAPIYIETWGTSDASGNGVFLDQSRWELVYLSKINISWNWLSDKQESSALKP 305
QY 265 --DLSNDMAGNNLTWAGNTIYNGPYGLRTSR 294
Db 306 GASKTGWPLTIDLASGTFRVRENIRGTKDSK 337

RESULT 12
GUN2_BACSU STANDARD; PRT; 499 AA.
ID GUN2_BACSU
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMC4SE) (CELLULOSE).
GN BGIC OR GLD OR EGTS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PAP115;
RC MEDLINE: 87066783.
RA Mackay R.M.; Lo A.; Willick G.; Zuker M.; Baird S.; Dove M.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.

```

RC STRAIN-CK-2;
 RA MEDLINE; 95225656
 RA Llandahl V., Aa K., Tronsmo A.;
 RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus*
 RT subtilis CK-2.";
 RL Antonie Van Leeuwenhoek 66:327-332(1994);
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Rose M., Enliar K.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE OF 30-45.
 RC STRAIN-CK-2;
 RA MEDLINE; 95225655.
 RA Aa K., Flensburg R., Llandahl V., Tronsmo A.;
 RT "Characterization of production and enzyme properties of an
 RT endo-beta-1,4-glucanase from *Bacillus subtilis* CK-2 isolated from
 RT compost soil.";
 RL Antonie Van Leeuwenhoek 66:319-326(1994).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Z29076; CAA82317.1; -;
 DR EMBL; X04689; CAA28392.1; -;
 DR EMBL; X67044; CAA47429.1; -;
 DR EMBL; Z73234; CAA97610.1; ALT_INIT.
 DR EMBL; Z99113; CAA13696.1; ALT_INIT.
 DR PIR; A26114; A26114.
 DR HSSP; Q06851; INBC.
 DR SUBTILIST; BG10437; BGIC.
 DR INTERPRO; IPR001547; -;
 DR INTERPRO; IPR001956; -;
 DR PFAM; PF00942; CBD_3; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR CELLULOSE degradation; Hydrolase; Glycosidase; signal.
 KW CHAIN 1 29
 FT SIGNAL 30 499
 FT ACT_SITE 169 169
 FT ACT_SITE 257 257
 FT DOMAIN 350 499
 FT CONFLICT 283 283
 SQ SEQUENCE 499 AA; 55287 MW; 8F735FE711B3EAE2 CRC64;
 Query Match 10.0%; Score 162.5; DB 1; Length 499;
 Best Local Similarity 21.0%; Pred. No. 2.2e-05;
 Matches 66; Conservative 75; Mismatches 136; Indels 37; Gaps 15;
 QY 3 NSGFGVSGTLLYDANGNPFVARGI-NHGHWYKQDQAT-TAIEGIANTGANTV---RIVLS 57
 DB 39 NGQSLSTGQTVNRDGNVQKSGISHLQWGYGVNKKDSLKWRDMDGIVFRFAAYTA 98
 QY 58 DGGGWTQDDT-HYVRNLSTLAEDNHLVAVPEVHDATGYDSIASNRAVDWIMERSALIG 116
 DB 99 DGGYIDNPSVKNKYKAEVAKELGIYVLLIDWHLNDGNPNQNEKKEFFKEM-SSLYG 157
 QY 117 KEDVYIINIAEWFG--SMEDG--AMADGYKQAIPLRNAGINHTLWADAAGMGQFQSI 172
 DB 158 NTPNVIYIETAEPPNDGVNWKDKIRPAE---EVSIVIRKNDPDIITV---GTSWQDV 211
 QY 173 HDYGRVFNADPO---RNTMFSIHMYEYAGNAGSOVRTNIDRVINQDLALVIGFGRHRT 229

DB 212 ND-----AADQLDANVMTALHF--YAGTHGQFLRKANVATLSKGAIPVTEMGSDA 263
 QY 230 NGD-----VDEA-TINSYBORGCVGLAWSKNGEWSEYLL---DLSNDWAGNNULIANGN 280
 DB 264 SGNQGVFLDQSHREMLKIDSKRTISVWNWNLSDPKRSSALRFGASKTGWMRLSDLSASGT 323
 QY 281 TIVNGPYGLRPTSR 294
 DB 324 EVRENILGTKDSSTK 337
 RESULT 13
 GUNA_STRLI STANDARD; PRT; 459 AA.
 ID GUNA_STRLI
 AC P27035;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE ENDOGLUCANASE CELA PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CELLULOSE).
 GN CELA.
 OS Streptomyces lividans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomyces; Streptomyces.
 CC [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-43.
 RC STRAIN-66;
 RX MEDLINE; 92246492.
 RA DeBerge M., Lacaze P., Shareck F., Morosoli R., Kluepfel D.;
 RT "Purification and characterization of an endoglucanase from
 RT Streptomyces lividans 66 and DNA sequence of the gene.";
 RT Appl. Environ. Microbiol. 58:815-820(1992).
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
 CC SITE FOR PROTEOLYSIS.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M82807; -; NOT_ANNOTATED_CDS.
 DR HSSP; P07986; LEXH.
 DR INTERPRO; IPR001547; -;
 DR INTERPRO; IPR001919; -;
 DR PFAM; PF00553; CBD_2; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00651; CBD_BACTERIAL; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR CELLULOSE degradation; Hydrolase; Glycosidase; signal.
 KW CHAIN 1 27
 FT SIGNAL 28 459
 FT DOMAIN 136 135
 FT DOMAIN 147 147
 FT DOMAIN 148 357
 FT ACT_SITE 286 286
 FT ACT_SITE 378 378
 FT DISULFID 31 131
 SQ SEQUENCE 459 AA; 48663 MW; CFF47EC34E297A7 CRC64;
 Query Match 9.9%; Score 161; DB 1; Length 459;
 Best Local Similarity 25.0%; Pred. No. 2.6e-05;
 Matches 81; Conservative 42; Mismatches 145; Indels 56; Gaps 16;

QY 3 NSGFVSGSTLTLYDANGNPFVNRGIN-HGNAM-----YKQATATAIGSANTGANTRYIVLS 57
 Db 156 NGQLHVCVGHLCNOYDRPQLNGKSTHGLQMGPCPTGDSLDRLQADMKSKDLLRYAMAYQ 215
 QY 58 DGGQWTKDDIHT--VRNLTSLAEDNHLVAVPVVDATGDSIASINRAVDYIEWRSALI 115
 Db 216 EDGYETDPAGFTSRVNGLYDMAEDGKMAVVIDPHLLTPEDPRVYNIDPRATFF-----SSVA 271
 QY 116 GKED--YTIINIANMFWG-SWEG-DAMADGYIQAIPLRNAGLNTLMAVDAKGQOFPOS 171
 Db 272 ARNDKKNVYIEIANPNPNGSVTAIVSYAE---QVLPVIRAAPDAVVIYGTGWSLSL-G 326
 QY 172 IHDYRE---VENADPQRNFMESIMHYEAGSAGNASQVPRNIDRYVNOGLALYIGSGRHH 228
 Db 327 VSDGANESEVVPVNPVATIMTAHFH--YAASHKDDYRAAV-RPATRIPLTFVSEFG--- 380
 QY 229 TNGDYDEATIMSYSBORGVGLAW--SWKGNPEWEYLDL-----SNDWAGN 273
 Db 381 ----TVSATFMSVDRSSSVAMTDLDDQLKISTANTYTSADRGSAAFRCGTCEGDYSS 436
 QY 274 NLTANGTIVNGPYGLRETSRST 297
 Db 437 GVLTEGALYK-----SRIST 452

RESULT 14
 GUN_BACSI
 ID GUN_BACSI STANDARD: PRT: 800 AA.
 AC P06564;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1993 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (ALKALINE CELLULOSE).
 OS Bacillus sp. (strain 1139).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87085443.
 RA Fukumori F., Kudo T., Narahashi Y., Horikoshi K.;
 RT "Molecular cloning and nucleotide sequence of the alkaline cellulase
 gene from the alkalophilic Bacillus sp. strain 1139.";
 RL J. Gen. Microbiol. 132:2329-2335(1986).
 CC -1- CATALYTIC ACTIVITY: ENDODIOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -1- MISCELLANEOUS: ALKALOPHILIC BACILLUS SP., STRAIN 1139, PRODUCES
 ONE ALKALINE CELLULOSE (PH OPTIMUM 9.0). THIS STRAIN IS NOT A TRUE
 CELLULOTYIC MICRO-ORGANISM BECAUSE THE ENZYME IS UNABLE TO
 HYDROLYSE NATIVE CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL: M15743; AAA22305.1; -;
 DR EMBL: D00066; BAA0045.1; -;
 DR PIR: A29003; A29003.
 DR INTERPRO: IPR001547; -;
 DR PFMAM: PF00150; cellulase, 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5, 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 800
 FT ACT_SITE 190 190
 FT ACT_SITE 305 305
 FT ACT_SITE 305 305
 FT NUCLEOPHILE (BY SIMILARITY).
 FT PROTON DONOR (BY SIMILARITY).
 FT CATALYTIC (BY SIMILARITY).

```

SQ      SEQUENCE    800 AA;   88602 MW;   70CAAD17B6DAD55CF CMC64;

Query Match          9.3%; Score 152; DB 1; Length 800;
Best Local Similarity 22.8%; Pred.No. 0.0026;
Matches 74; Conservative 53; Mismatches 125; Indels 72; Gaps 18;

QY      12 TLYDANGNPFWRGIN-HGHANYKQOAT-TAIEGIAIT-GANTVAIVL--SDGGCWTQDD 66
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       67 TLVDHGERGIQRIGRSTGLQWPEILNDNMYKALANWESNMRLTAAVSENGASNP 126
QY      67 IHTVANL--ISLAENHLVAAPVEH-DATGDSIASLARAVDYIEMSAIIGKEDYII 123
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       127 LTKSVYIGIDIALINDMVIYDVHVHAPGDPRPYVAGADEFFDI-AAALPNPHITY 185
QY      124 NIANWFSGWEGDA-----WADGYKOATP---RLNNAGL-NHILMDAAGWGQFPQ 170
        :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       186 ELANPFSSNNNGACAIIPNNEBGMAVKRYADIYELMLDSGNADNIILIVGSPNNQSOPD 245
QY      171 SIHDYGREVENADPORNTWESIHT-----ETAGGNASQVRNIDRYANDL 217
        -|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|
DB       246 LAAD-----NFIDDHMTYTVHYFTGSHAASTESPYPETPNSESGNYSMTRYALENG 299
QY      QY      218 ALVIEFPGHRHNGD----VDPAT-MAYSQDRGWLANE-----WKG 256
        :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       300 AFAFATEMGTSQANGSGGFPEADADWIEFLNENNISMAMSLTNKNEYSGAFTPPELGKS 359
QY      QY      257 N-----GPE-W--EYLDLSNDM 270
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       360 NATSLDPGPDOYWPPELSTSGEY 383

RESULT 15
GN3_BACSU ID GN3_BACSU STANDARD; PRT: 499 AA.

AC P23549; 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE ENOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
CN BGIC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BSE616;
RX MEDLINE; 91299280.
RA Park S.H., Kim H.K., Pack M.Y.;
RA "Characterization and structure of the cellulase gene of Bacillus
RT subtilis BSE616.";
RU Agric. Biol. Chem. 55:441-448(1991).
CC -1- CATALYTIC ACTIVITY: ENDOMYOHLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D01057; BAA00859.1; -.
DR PIR; JN0111; JN0111.
DR HSSP; O06851; INBC.
DR INTERPRO; IPRO001547; -.
DR INTERPRO; IPRO001956; -.
DR PFAM; PF00942; CBD_3; 1.
DR PFAM; PF00150; cellulase; 1.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:06:42 ; Search time 248.55 Seconds
(without alignments)
76.597 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGFYVSGTITLYDANGNP.....TIVNGPYGLAETSRSLSTVFF 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_65:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1169	71.7	516	2 JE0134	mannan endo-1,4-be
2	710	43.6	384	2 T37232	secreted beta-mann
3	692	42.5	363	2 S30386	mannan endo-1,4-be
4	601	36.9	1331	2 A48954	mannan endo-1,4-be
5	202	12.4	409	2 B25156	cellulase (EC 3.2.
6	200	12.3	488	2 A25156	cellulase (EC 3.2.
7	197	12.1	557	2 JC5487	cellulase (EC 3.2.
8	184.5	11.3	505	2 S39662	endoglucanase - Er
9	180	11.0	448	2 A27631	cellulase (EC 3.2.
10	176.5	10.8	466	2 C42360	cellulase (EC 3.2.
11	174.5	10.7	499	2 A27198	cellulase (EC 3.2.
12	174.5	10.7	504	2 S54744	cellulase (EC 3.2.
13	173.5	10.6	508	2 A26874	cellulase (EC 3.2.
14	162.5	10.0	486	2 I40548	bifunctional cellu
15	162.5	10.0	508	2 G69593	cellulase (EC 3.2.
16	157	9.6	783	2 JC5467	cellulase (EC 3.2.
17	152	9.3	800	2 A29003	cellulase (EC 3.2.
18	152	9.3	822	2 JT0611	cellulase (EC 3.2.
19	151.5	9.3	499	2 JN0111	cellulase (EC 3.2.
20	151.5	9.3	825	2 JN0174	cellulase (EC 3.2.
21	149.5	9.2	357	2 PC4404	cellulase (EC 3.2.
22	149.5	9.2	941	2 S29043	cellulase (EC 3.2.
23	147.5	9.0	429	2 S29044	endoglucanase A pr
24	136.5	8.4	570	2 S56132	cellulase (EC 3.2.
25	131	8.0	814	1 C2CLEM	cellulase (EC 3.2.
26	130	8.0	426	2 A42649	cellulase (EC 3.2.
27	127	7.8	32	2 PC4278	guar gum-degrading
28	122.5	7.5	915	2 A43802	cellulase (EC 3.2.
29	122.5	7.5	1039	2 S02711	cellulase (EC 3.2.

30	122	7.5	438	2 A47702	glucan 1,3-beta-gl
31	117.5	7.2	428	2 S03767	cellulase (EC 3.2.
32	116	7.1	611	2 JC7177	endoglucanase V (E
33	115.5	7.1	747	2 B47093	cellulase (EC 3.2.
34	114	7.0	419	2 S72325	glucan 1,3-beta-gl
35	113.5	7.0	517	2 I40798	cellulase (EC 3.2.
36	111.5	6.8	441	2 A44815	cellulase (EC 3.2.
37	111	6.8	475	1 C2C1CA	cellulase (EC 3.2.
38	110	6.7	26	2 PC4285	guar gum-degrading
39	109.5	6.7	754	2 T14877	hypothetical prote
40	107	6.6	430	2 S55325	endo-beta-1,6-gluc
41	105	6.4	364	2 S12017	endoglucanase A -
42	105	6.4	456	2 T40276	probable gaal faml
43	104	6.4	566	2 A40589	cellulase (EC 3.2.
44	101.5	6.2	317	2 B72216	endoglucanase - Th
45	100.5	6.2	748	2 S19652	cellodextrinase C

ALIGNMENTS

RESULT 1
JE0134
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N:Alternate names: endo-1,4-beta-mannanase
C:Species: Bacillus circulans
C:Date: 03-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C:Accession: JE0134
R:Yoshida, S.; Sako, Y.; Uchida, A.
BioSci. Biotechnol. Biochem. 62, 514-520, 1998
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: JE0134; MUID:98233274
A:Accession: JE0134
A:Molecule type: mRNA
A:Residues: 1-516 <YOS>
A:Cross-references: DDBJ:AB007123
A>Note: the authors translated the codon CAA for residue 259 as Asn, CAA for residue
A>Note: the translation of the nucleotide sequence 1294-1365 is not given in this pap
C:Keywords: glycoprotein; glycosidase; hydrolase

Query Match	Best Local Similarity	Score	Pred. No.	Length	DB 2:	Length	DB 2:
Matches	207	Conservative	48	Mismatches	44	Indels	0
Gaps	0						
QY	1	NANSGFYVSGTITLYDANGNP	71.7%	516	2	JE0134	mannan endo-1,4-be
DB	33	HAASGFYVSGTITLYDANGNP	69.2%	384	2	T37232	secreted beta-mann
QY	61	QNTKDIHTVKNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWTMRGALIGKEDT	48	409	2	B25156	cellulase (EC 3.2.
DB	93	KWTLDDVNTVNNITLCEQNKRIAVLEVHDATGSDLSLDNAVYWGIRKALIGKEDR	42.5	363	2	S30386	mannan endo-1,4-be
QY	121	VIIINIANEWFSGWEDDANADGKOAIPRLRNGLVHTLMVDAAGGQPPQSHDGRFV	36.9	1331	2	A48954	mannan endo-1,4-be
DB	153	VIIINIANEWFSGWEDDANADGKOAIPRLRNGLVHTLMVDAAGGQPPQSHDGRFV	36.9	1331	2	A48954	mannan endo-1,4-be
QY	181	NADPQNTWFSIMTEYAGNAGSOVRTNIDRYLNDLALVIEFGRHNGVDYDATTIS	12.4	409	2	B25156	cellulase (EC 3.2.
DB	213	NADPQNTWFSIMTEYAGNAGSOVRTNIDRYLNDLALVIEFGRHNGVDYDATTIS	12.4	409	2	B25156	cellulase (EC 3.2.
QY	241	YSEQKGVGLWANSKMGNGPEWEDLSDNMGANNLTAMGNTIVNGPYGLRSTSLSTVF	11.3	505	2	S39662	endoglucanase - Er
DB	273	YSEQKGVGLWANSKMGNGPEWEDLSDNMGANNLTAMGNTIVNGPYGLRSTSLSTVF	11.3	505	2	S39662	endoglucanase - Er
RESULT 2							
T37232							
secreted beta-mannosidase - Streptomyces coelicolor							
C:Species: Streptomyces coelicolor							
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999							
C:Accession: T37232							
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.							

submitted to the EMBL Data Library, September 1998

A:Reference number: 221615
A:Accession: T37232
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-364
A:Cross-references: EMBL:AL031514; PIDN:CAA20610.1; GSPDB:GNO0070; SCOEDB:SC2H4.16
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: manA2; SCOEDB:SC2H4.16

Query Match 43.6%; Score 710; DB 2; Length 384;
Best Local Similarity 46.7%; Pred. No. 4,4e-48;
Matches 141; Conservative 53; Mismatches 102; Indels 6; Gaps 4;

QY 2 AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIYSDGQ 61
DB 37 AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIYSDGQ 61
QY 62 WTAKDIIHTVRLNLSLAEDNHLVAPEVHDATGY--DSIASLRAYDIEMRSALIGKE 118
DB 96 WTAKDIIHTVRLNLSLAEDNHLVAPEVHDATGY--DSIASLRAYDIEMRSALIGKE 118
QY 119 DTVIINIANEMGSGWEGDAMADYKQALPRLNAGLNHTLMDVADAGMGQ-FPOSIDYGR 177
DB 156 DTVIINIANEMGSGWEGDAMADYKQALPRLNAGLNHTLMDVADAGMGQ-FPOSIDYGR 177
QY 178 EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGDVDEAT 237
DB 216 EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGDVDEAT 237
QY 238 IMSYBQRCVGLWASWKGNGPEWEYLDLSNDWAGNNLTAMGNTVNGPYGLRETSRLST 297
DB 275 IMATQSLGVGYLGWSWNGGVEYLDLVNGDFDPSNLTSMGNTIRYSGNGIAATSRTAT 334
QY 298 VF 299
DB 335 VF 336

RESULT 3
S30386
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) precursor - Streptomyces lividans
N:Alternate names: beta-mannanase
C:Species: Streptomyces lividans
C:Date: 02-Dec-1993 #sequence_revision 13-Mar-1997 #text_change 22-Oct-1999
A:Accession: S30386; S27699
R:Arcand, N.; Kuepfel, D.; Paradis, F.W.; Morosoli, R.; Shareck, F.
Biochem. J. 290, 857-863, 1993
A:Title: beta-Mannanase of Streptomyces lividans 66: cloning and DNA sequence of the man
A:Reference number: S30386; MUID:93207541
A:Accession: S30386
A:Molecule type: DNA
A:Residues: 1-363 <ARC>
A:Cross-references: EMBL:M92297; NID:g153193; PIDN:AAA26710.1; PID:g404076
A:Experimental source: strain 1326
C:Genetics:
A:Gene: manA
C:Function:
A:Description: catalyses degradation of beta-mannans into short-chain oligo-mannosides;
A:Pathway: mannose metabolism
A>Note: Industrially used for bioleaching of wood pulps
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <sig>
F:37-363/Product: mannan endo-1,4-beta-mannosidase #status predicted <mat>

Query Match 42.5%; Score 692; DB 2; Length 363;
Best Local Similarity 46.7%; Pred. No. 1e-46;
Matches 141; Conservative 51; Mismatches 104; Indels 6; Gaps 5;
QY 2 AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIYSDGQ 61

DB 36 AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIYSDGQ 94
QY 62 WTAKDIIHTVRLNLSLAEDNHLVAPEVHDATGY--DSIASLRAYDIEMRSAL-IGKE 118
DB 96 WTAKDIIHTVRLNLSLAEDNHLVAPEVHDATGY--DSIASLRAYDIEMRSAL-IGKE 118
QY 119 DTVIINIANEMGSGWEGDAMADYKQALPRLNAGLNHTLMDVADAGMGQ-FPOSIDYGR 177
DB 156 DTVIINIANEMGSGWEGDAMADYKQALPRLNAGLNHTLMDVADAGMGQ-FPOSIDYGR 177
QY 178 EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGDVDEAT 237
DB 216 EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGDVDEAT 237
QY 238 IMSYBQRCVGLWASWKGNGPEWEYLDLSNDWAGNNLTAMGNTVNGPYGLRETSRLST 297
DB 275 IMATQSLGVGYLGWSWNGGVEYLDLVNGDFDPSNLTSMGNTIRYSGNGIAATSRTAT 334
QY 298 VF 299
DB 334 VF 335

RESULT 4
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
A:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multimer
A:Reference number: A48954; MUID:93119139
A:Accession: A48954
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
A>Note: sequence extracted from NCBI backbone (NCBI:121576, NCBI:121577)
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cc
A:Reference number: A43745; MUID:91247819
A:Accession: B43745
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PROHORO' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A>Note: the authors translated the codon CAC for residue 262 as Glu
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.9%; Score 601; DB 2; Length 1331;
Best Local Similarity 44.5%; Pred. No. 7,3e-39;
Matches 125; Conservative 42; Mismatches 108; Indels 6; Gaps 4;
QY 25 GINHGAMTKDQATTAIEGANTGANTVRIYSDGQNTKDDIHTVRLNLSLAEDNHLVA 84
DB 55 GINHGAMTKDQATTAIEGANTGANTVRIYSDGQNTKDDIHTVRLNLSLAEDNHLVA 84
QY 85 -VPEVHDATGY--DSIASLRAYDIEMRSALIGKEDYITINIANEMGSGWEGDAMAD 140
DB 115 ILEVHDATGY--DSIASLRAYDIEMRSALIGKEDYITINIANEMGSGWEGDAMAD 140
QY 141 GYKQALPRLNAGLNHTLMDVADAGMGQ-FPOSIDYGRFVNADPQNTMFSIMHEYAG 199
DB 175 GYKQALPRLNAGLNHTLMDVADAGMGQ-FPOSIDYGRFVNADPQNTMFSIMHEYAG 199
QY 200 GNASGVRTNIDRYLNODLALVIGFGRHRTNGDVDEATIMSYSBQRCVGLWASWKGNGP 259
DB 233 GNASGVRTNIDRYLNODLALVIGFGRHRTNGDVDEATIMSYSBQRCVGLWASWKGNGP 259

Db 234 NTASKVEEYIKSFVQKGLPVIYGFHGHQHTDGDDEALIVAKQYKIGLSMSMGSS 293
OY 260 EMEYLDLSNDMAGNNLTAMGNTIVNGPYGLAETSRSTVFT 300
Db 294 YVGTLDMVNNNDPNNPMPGQWYKTNALGISSTPTSTVFT 334

RESULT 5

B25156
cellulase (EC 3.2.1.4) 2 - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
R:Fukunori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic Bacillus sp. str.
A:Reference number: A91825; MUID:87056924
A:Accession: B25156
A:Molecule type: DNA
A:Residues: 1-409 <PUK>
A:Cross-references: GB:M14729; NID:q142655; PIDN:AAA22299.1; PID:q142656
A:Experimental source: strain N-4, plasmid pNK2
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 12.4%; Score 202; DB 2; Length 409;
Best Local Similarity 25.0%; Pred. No. 2.1e-08;

Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

OY 8 VSGTLLDANGNPFVPMGT-NHGAWTKQATTAIEGANTGANTVRIYSDGQWTRKD 66
Db 40 ISNGELVNDGEPYQKGMSSHGLOWYG-----QFVNYESMKWLRDD 81
OY 67 -----IHT-----VRNLISLAEDNHLVAVPEVHATGDSIASLNR 103
Db 82 WGITVFRANWTSSGGYIEDPSVKEKVEAVEALDGIYIIDMHLISDMPITYKEA 141
OY 104 VDIYIEKRSALIGKEDVTIINANEMFGS---WEGD---AMADGYKQAIPLRNAGNHTL 158
Db 142 KDFEDEN-SELYGDPYVNIYIEIANEPNSDVTWNOIKPYAE---EYIYVIRNNDPNNIT 197
OY 159 MVDAAAGGQFPGSIHDYGREVFNADPQNTMFSIHMYETAGNAGSQQVNTIDRYLNDOLA 218
Db 198 IV---GTGWSQDVH-HAADNQLDP--NVMYAFHF--YAGTHGQNLADQVYDALDQGA 249
OY 219 LVIGEFGRHRTNGD---VDEATI-MSYSBOGCVGLAWS 253
Db 250 IFVSEWGTSEATGDGVFLDEAQWIDFMDERNLSNANWS 289

RESULT 6

A25156
cellulase (EC 3.2.1.4) 1 - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
R:Fukunori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic Bacillus sp. str.
A:Reference number: A91825; MUID:87056924
A:Accession: A25156
A:Molecule type: DNA
A:Residues: 1-488 <PUK>
A:Cross-references: GB:M14781; GB:X53449; NID:q142659; PIDN:AAA22301.1; PID:q142660
A:Experimental source: strain N-4, plasmid pNK1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 12.3%; Score 200; DB 2; Length 488;

Best Local Similarity 24.6%; Pred. No. 3.8e-08;

Matches 70; Conservative 49; Mismatches 103; Indels 62; Gaps 14;

OY 3 NSGFYVSGTLLDANGNPFVPMGT-NHGAWTKQATTAIEGANTGANTVRIYLSGCG 61
Db 33 NGQLSINGQLVNEHGPVQLKGMSSHGLOWYG-----QFVNYYSIK 74
OY 62 WTKDD-----IHT-----VRNLISLAEDNHLVAVPEVHATGYSIA 98
Db 75 WLKDDKGIYFRAMYTSSGGYIEDPSVKEKVEAVEALDGIYIIDMHLISDNDPNI 134
OY 99 SLNDAVDYIEKRSALIGKEDVTIINANEMFG---SWEGDAMADGY-KQAIPLRNAGL 154
Db 135 YKEBAKFEFDEM-SALYGDYVNIYIEIANEPNGHNVM--DSHIKPYAEVYIPVIRANDP 191
OY 155 NHTLMVDAAGGQFPGSIHDYGREVFNADPQNTMFSIHMYETAGNAGSQQVNTIDRYL 214
Db 192 NNIVYGTATWS---QVHEAADNQLD-DP--NVMYAFHF--YAGTHGQQLRNQVYALS 243
OY 215 QDLALVIGEFGRHRTNGD---VDEATI-MSYSBOGCVGLAWS 253
Db 244 RGAALIFVSEWGTSEATGDGVFLDEAQWIDFMDERNLSNANWS 287

RESULT 7

JC5487
cellulase (EC 3.2.1.4) precursor - Clostridium cellulovorans
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium cellulovorans
C:Date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 15-Oct-1999
C:Accession: JC5487; PC4333
R:Shewetta, S.A.; Ichii-Ishii, A.; Park, J.S.; Liu, C.; Malburg, L.M.; Dol, R.H.
Gene 182, 163-167, 1996
A:Title: Characterization of engF, a gene for a non-cellulosomal Clostridium cellulov
A:Reference number: JC5487; MUID:9715706
A:Accession: JC5487
A:Molecule type: DNA
A:Residues: 1-557 <SHE1>
A:Cross-references: GB:U37056; NID:q1778708; PIDN:AA840891.1; PID:q1778709
A:Experimental source: strain ATCC 35296
A:Accession: PC4333
A:Molecule type: protein
A:Residues: 30-37 <SHE2>
C:Comment: This enzyme plays a secondary role in cellulose degradation.
C:Genetics:
A:Gene: engF
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-557/Product: cellulase #status predicted <MAV>

Query Match 12.1%; Score 197; DB 2; Length 557;

Best Local Similarity 26.7%; Pred. No. 7.7e-08;

Matches 88; Conservative 33; Mismatches 132; Indels 76; Gaps 17;

OY 12 TLVDANGNPFVPMGT-NHGAWTKQATTAIEGANTGANTVRIYLS-SD 58
Db 62 TLCDKDNPIQLKGMSSHGLOWY-----PGVYNNAPALASNDMSNVIRIATYAE 113
OY 59 GGQWTKNDI-HTVRNLISLAEDNHLVAVPEVHATGDSIASINRAVDYIEKRSALIGK 117
Db 114 GGVATNPVSQKQYINGINVTALNDMTYIVDMHMMNPDPNAAVYSGQSFFNISTILY 173
OY 118 EDYIYIINANEMFGS---DA--WADGYKQAIPL---KRNAGNHTLMVDAAGGQF 169
Db 174 NKNIIYELCNEPENGEGVNDATGMAQVRSYATPIYQLLRNNGENLITVGNPFSQRP 233
OY 170 QSIHDYGREVFNADPQNTMFSIHMYETAGN-ASQVNTIDR-----VLQDLALVI 221

Db 234 DLAAD-----NPINDSNTMTSVH--YSGNPISITVDTPNRDNMSNVRYALNNGHAAVFA 285
QY 222 GEGHRTNGDVDEATIMSISEQRCVGLAWSKNGPEWEYLDSND----- 269
Db 286 TEMGTSIATG-----TTPPYL-ARKADAWLDF-LNGNINISMCNFSISNDEKRAAALNLSITS 338
QY 270 -----MAGNNLWAGNTI---VNGPY 287
Db 339 LDPSGDKMLADNELITTSQYVRAKKGAY 367

RESULT 8

339962
endoglucanase - Erwinia carotovora

C:Species: Erwinia carotovora
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39962
R:Cooper, V.J.C.; Salmond, G.P.C.
Mol. Gen. Genet. 241, 341-350, 1993
A:Title: Molecular analysis of the major cellulase (CelV) of Erwinia carotovora: evidenc
A:Reference number: S39962; MUID:94067016
A:Accession: S39962
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-505 <COO>
A:Cross-references: EMBL:X76000; NID:9434941; PIDN:CAA53592.1; PID:9434942

Query Match 11.3%; Score 184.5; DB 2; Length 505;
Best Local Similarity 23.1%; Pred. No. 6.4e-07;
Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

QY 13 LYDANGNFVNRGI-NHGHAWYKQATTAIGSINTGANTYRIYLSDDGQTKPD----- 66
Db 48 LVDEGKRVQLNGISSRGQLWFGD-----YVNRKDSMKVLRODMGINTV 89
QY 67 -----IHTVRNLISLAEDNHLVAPEVDATGYDSIASLNRADVWI 108
Db 90 FRVAMTAAADGYISNPSLANVKEVAAQSIGYIIITDHLISNDNPITKAAQKTFFA 149
QY 109 EMRSALICKEDTYIINANWFG--SWEGD--AWADGYKQAIPLRNAGLHNTLWVDAAG 164
Db 150 EM-AGLYSSPNVYIEINENPGGVTWNGQIRPYA---LEVTDTRSKDPNLTIV---G 202
QY 165 WGPQSHDYGREYFNADPQRTMESIHMYEYAGNAGSQRINIDRYLNDLALVIGEF 224
Db 203 TGTWSQDIDH--AADNOLPDPNTYALHF--YAGTGHQFLRDRIDVAGSRGALFVSEW 257
QY 225 GHRHTNGD-----VDEATIMSISEQRCVGLAWSKNGPEWEYL---DLSNDMAGNLT 275
Db 258 GTSASGNGGFLPESQTIWIDFINNRGYSWNWBSLTDKSEASALAPGASKSGGTQNL 317
QY 276 TANGNTI 282
Db 318 STSGKTV 324

RESULT 9

A27631

cellulase (EC 3.2.1.4) precursor - Clostridium acetobutylicum

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Clostridium acetobutylicum

C:Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999

C:Accession: A27631

R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.

Appl. Environ. Microbiol. 54, 1289-1292, 1988

A:Title: Structure of an endo-beta-1,4-glucanase gene from Clostridium acetobutylicum P2

A:Reference number: A27631; MUID:88268074

A:Accession: A27631

A:Molecule type: DNA

A:Residues: 1-448 <ZAP>

A:Cross-references: EMBL:M31311; NID:9144789; PIDN:AAA23230.1; PID:9144790

A:Note: the authors translated the codon GAG for residue 116 as Gly, GAA for residue 263

C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 11.0%; Score 180; DB 2; Length 448;
Best Local Similarity 22.1%; Pred. No. 1.2e-06;
Matches 71; Conservative 59; Mismatches 135; Indels 56; Gaps 14;

QY 8 VSGTTLTDANGNPFVNRGI-NHGHAWY-----KQATTAIGSINTGANTVRI 54
Db 49 VVGSQLCDNSGNKPIQLKGMSSHGLWYVNFVNDKMKFLDK-----MGVNVIRA 98
QY 55 VL--SDGGQNT--KDIHTVRLNLSLAEDNHLVAPEVDHATGYDSIASLNRADVWTEW 110
Db 99 AMYINTEGYSINSSQKREKRIYQDAIDNMVYIIDHILSNPNPTKYEQAKSFEQEM 158
QY 111 RSALIGKEDTYIINANWFGSWEGDADGVR---QAIPLRNAGLHNTLWVDAAGW 166
Db 159 AEF-YGKISNVYIEICNEPNG--GYNWADIKFYANTYIIPALRAIDPNITIVGTSWS 214
QY 167 QPQSHDYGREYFNADPQRTMESIHMYEYAGNAGSQRINIDRYLNDLALVIGEF 224
Db 215 QDVVIDAD-----NPLRYSNIMYTCHE--YAGTHTQSLRDKINAMSKGIAIFVTEW 264
QY 225 GHRHTNGD-----VDEA-TIMSISEQRCVGLAWSKNGPEWEYL---DLSNDMAGNLT 275
Db 265 GTSASGNGGFLPESQTIWIDFINNRGYSWNWBSLTDKSEASALAPGASKSGGTQNL 324
QY 276 TANGNTIYNGPYGLRSTRSL 296
Db 325 TTSGLFYKKSIGSNTTSQTS 345

RESULT 10

C42360

cellulase (EC 3.2.1.4) E5 precursor - Thermomonospora fusca

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Thermomonospora fusca

C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 20-Mar-1998

C:Accession: C42360

R:Iao, G.; Ghanges, G.S.; Jung, E.D.; Wilson, D.B.

J. Bacteriol. 173, 3397-3407, 1991

A:Title: DNA sequences of three beta-1,4-endoglucanase genes from Thermomonospora fus

A:Reference number: A42360; MUID:91258320

A:Accession: C42360

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-466 <LAO>

A:Cross-references: GB:I01577; NID:9154693; PID:9154694

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: bacterial cellulose-binding domain homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:38-137/Domain: bacterial cellulose-binding domain homology <CB>

Query Match 10.8%; Score 176.5; DB 2; Length 466;
Best Local Similarity 25.0%; Pred. No. 2.4e-06;
Matches 78; Conservative 48; Mismatches 125; Indels 61; Gaps 17;

QY 8 VSGTTLTDANGNPFVNRGIN-HGHAWYKQAT-TAIEGIA-NTGANTVRI--VLSDDGQW 62
Db 173 VCGTQLCDEHNGVQLRGSMSTHGIOWFDHCLTDSIDALAYDWKADIRLRSMYIOEDGYE 232
QY 63 TKQDIIHVR--NLISLAEDNHLVAPEVDHATGYDSIASLNRADVWTEWRSALIKEDT 120
Db 233 TNRGFTDRIDQLDADATAGLVYVDMWHLTPGDPHYNIDRAKTFEAFEAQRAHAKTN- 291
QY 121 VIINANWFGSWEGDAMD--GY-KQAIPLRNAGLHNTLWVDAAGWQFQPSIDHYGR 177

```
Db      292 VLYEIANE-----PAGVSMASIKSYAEIEVIPIHQBDPSVIIYGTGWSLSLGS-EGSGP   346
Oy      178 EVFNADP--QRNINFSIHTEIYAGGNASQVRINDRYLNODLL--VIGESGHRTNSD   232
          :|::||::||::||::||::||::||::||::||::||::||::||::||
Db      347 AEIANEPVNASINIMYAFHFY-----AASRDNYLNALREASELFPEVFYEFTETYYTD   400
Oy      233 -----VDKATINSYEQROGVGLAMSKR-----GNGPMEWEIDLSNDW    270
          :|::||::||::||::||::||::||::||::||::||::||::||
Db      401 GANDFOMADR--YIDLMAERRIGTKWNYSDDFRSGAVFOPTCASGGP-----W    448
Oy      271 AGNNLTANGNTI   282
          :|::||| | | | |
Db      449 SGGSLTASGQNV   460
```

RESULT 11

A27198

cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain IF03034)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Bacillus subtilis*

C:date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999

C:Accession: A27198

R:Nakamura, A.; Tozumil, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987

A:Title: Nucleotide sequence of a cellulase gene of *Bacillus subtilis*.

A:Reference number: A27198; PMID:87190357

A:Accession: A27198

A:Molecule type: DNA

A:Residues: 1-499 <NAK>

A:Cross-references: GB:W28332; NID:g142670; PIDD:AAA22307.1; PID:g142671

A:Experimental source: strain IF03034

A:Description hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose
 A:Pathway: cellulose degradation
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match	10.7%	Score 174.5	DB 2	Length 459
Best Local Similarity	21.1%	Pred. No. 3.8e-06		
Matches 70; Conservative	66; Mismatches 123;		Indels 73;	Gaps 15

[illegible]

RESULT 12
S54744
cellulase (EC 3.2.1.4) CelY1 precursor - *Erwinia carotovora* (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: *Erwinia carotovora*
A:Variety: SCC 3193

C|date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C|Accession: S54744; S44996
R|Maez, A.; Heikinehmo, R.; Palva, E.T.
Mol. Gen. Genet. 247: 17-26, 1995
A|Title: Structure and regulation of the *Erythrina carotovora* subspecies *carotovora* S54744
A|Reference number: S54744; M01D:95231512
A|Accession: S54744
A|Molecule type: DNA
A|Residues: 1-504 <MAE>
A|Cross-references: EMBL:X79241; NID:G493492; PIDD:CAAS5823.1; PID:G493493
C|Genetics:
A|Gene: celV1
C|Function:
A|Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A|Pathway: cellulose degradation
C|Keywords: glycosidase; hydrolase; polyanthracide degradation
E|1-32/Domains: signal sequence #status predicted <SIG>
E|33-504/Product: cellulase #status predicted <CAT>

Query Match	Similarity	10.7%	Score 174.5	DB 2	Length 504
Best Local	Similarity	23.7%	Prod. No. 3.9e-06		
Matches	Conservative	43	Mismatches	123	Indels 69
			Gaps		
DB	13	LYDANGNPFVAKGI-NHGHWATKQOATTAEGIANGTANTVRIYVLDGGGWKDD--IHT	69		
	48	LYDEGGKRVKQVLRGSISSNQLQWGD-----YNNKOSMKWLDDMGVING	89		
QY	70	VANLISLAEDNHLVAPVY-----HDATGDSIASLNRADVIV	107		
DB	90	FVNAVYATFE-NGYIATNSPIANKVEAAVAAAGLGCVIIIDMHWLSDDDPPTYRAQAKIEF	148		
QY	108	IEMSRSLAGKEDPYIYNINANEFMS--WEGD--AMADYGKOAIPRLNAGLNTHTLVDA	163		
DB	149	AEH-AGLGNSPNVITIELANEPNCSYVMNQIRFYA---LEVYDTISKRDPDLIIYVGS	204		
QY	164	GWGQFPQSIHGYGREVFADPQRTMFSIHXYEYAGNASQAVOVRTNIDRYLMDLALVIGE	223		
DB	205	TWS---QDINH---AADNQLPDPNTLYALHF--YAGTHGQFLNRIDRYAQSRAAIFVSE	256		
QY	224	FGHRTND-----VDEATINSYEQGVGLMAMSKNGNPEWEYL-----DLSNDAGNN	274		
DB	257	WGTSDASNGGPFPLPESQWTIDFLNNRGISWVMSLSDKSETSAALVAGASKSGGWTEON	316		
QY	275	LTAMGNTI 282			
DB	317	ISTSGKFV 324			

RESULT 13

A26874

cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain DLG)

N.Alternate names: endo-1,4-beta-glucanase

C.Species: *Bacillus subtilis*

C.Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999

C.Accession: A26874; B26874

R.Robson, L.M.; Chambliss, G.H

J.Bacteriol. 169, 2017-2025, 1987

A.Title: Endo-beta-1-4-glucanase gene of *Bacillus subtilis* DLG.

A.Reference number: A26874; M01D:87194581

A.Accession: A26874

A.Molecule type: DNA

A.Residues: 1-508 <R0B1>

A.Cross-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008

A.Experimental source: strain DLG

A.Accession: B26874

A.Molecule type: protein

A.Residues: 39-53 <R0B2>

A.Experimental source: strain DLG

A.Note: the authors believe Met-1 and Met-2 may be alternate initiators

C.Comment: The low molecular weight of the mature protein suggests carboxyl-terminal

C.Function: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such

A:Pathway: cellulose degradation
 C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; polys
 E:1-38/Domain: (or 2-38) signal sequence #status predicted <Sig>

Query Match 10.6%; Score 173.5; DB 2; Length 508;
 Best Local Similarity 20.8%; Pred. No. 4.7e-06;
 Matches 69; Conservative 68; Mismatches 122; Indels 73; Gaps 15;

QY 3 NSGFYVSGTLLDANGNPFVNGI-NHGHAWYKQATTAIEGIANTGANTVAVIYSDGQ 61
 DB 48 NSGLSTKGTQLVNRDCKAVOLKGISSHGLQWGD-----FVNRDSLK 89
 QY 62 WKPKD-----IHT-----VRLSLAEDNHLVAVPEVHATGDSLA 98
 DB 90 WLBDGDTVFRKMTADGDTINDSVKKNKAEVAEAKELGITYIIDHILINDGNPNQ 149
 QY 99 SLNRADVYIEKRSALIGEDYIINIANEWFG--SWEGD--AMADGYKQAIPLRNAGL 154
 DB 150 NKEKAEFEKEM-SSLYGMTPNYIETIANEPNGDVWKRKDIPYAE--EYISVIRKNDP 205
 QY 155 NHTLWDAAGWGFOSIHDYGEVFNADPO--RNTMFSIMHTEYAGNNSOVRTNDR 211
 DB 206 DNIIV--GTGTWQSDVD-----AADDQLKDANVYALHF--YAGTHGQSLRDKANY 254
 QY 212 VLNODLALYIEGHRHTNGD-----VDEA-TIMSYSEORGVGLAMSKNGPEWEYL-- 264
 DB 255 ALSKAPVYTEKISDASNGSVFLDQREMINIYDSKNISVWNNLSDKOBSSALNP 314
 QY 265 --DLSDNAGNNLTJANGNTIVNGPYGLRETSR 294
 DB 315 GASKTGWPLTDLTASGTFRREINRGTDKSTK 346

RESULT 14
 140548
 biFunctional cellulase precursor - Bacillus sp.
 C:Species: Bacillus sp.
 C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Oct-1999
 C:Accession: 140548

R:Han, S.J.; Yoo, Y.J.; Kang, H.S.
 J. Biol. Chem. 270, 26012-26019, 1995
 A:Title: Characterization of a bifunctional cellulase and its structural gene: the cel
 A:Reference number: 140548; MUID:96029707
 A:Accession: 140548
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-486 <RES>
 A:Cross-references: EMBL:U27084; NID:9857575; PIDN:AA43478.1; PID:9857576
 C:Genetics:
 A:Gene: cel

Query Match 10.0%; Score 162.5; DB 2; Length 486;
 Best Local Similarity 21.0%; Pred. No. 3.2e-05;
 Matches 66; Conservative 75; Mismatches 136; Indels 37; Gaps 15;

QY 3 NSGFYVSGTLLDANGNPFVNGI-NHGHAWYKQATTAIEGIANTGANTVAVIYSDGQ 57
 DB 39 NSGLSTKGTQLVNRDCKAVOLKGISSHGLQWGD-----FVNRDSLK 98
 QY 58 DSGOWTKDDI-HVRLMLISLADNHLVAVPEVHATGYDSIASLNRAVDYIEKRSALIG 116
 DB 99 DSGYIDNPVKKKAEVAEAKELGITYIIDHILINDGNPNQKKEAKFEFFKEM-SSLIG 157
 QY 117 KEDYIINIANEWFG--SWEGD--AMADGYKQAIPLRNAGNHTLMDVAGWGFOSI 172
 DB 158 NTPNVIYETIANEPNGDVWKRKDIPYAE--EYISVIRKNDPNIIV--GTGTWQSDV 211
 QY 173 HDYGRVFNADPO--RNTMFSIMHTEYAGNNSOVRTNDRVNDLALVIEGHRHT 229
 DB 212 ND-----AADDQLKDANVYALHF--YAGTHGQFLRDKRANVYALSKGAPVTEWGTSDA 263

QY 230 NGD-----VDEA-TIMSYSEORGVGLAMSKNGPEWEYL-----DLSDNAGNNLTJANGN 280
 DB 264 SNGSVFLDQSEWKLKIDSKTISVWNNLSDKOBSSALNPAGSKTGWRLSDLSASGT 323
 QY 281 TIVNGPYGLRETSR 294
 DB 324 FVRENIIGTKDSTK 337

RESULT 15
 669593
 cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text change 15-Oct-1999
 C:Accession: 669593; A26114; I40353; S24239; S49103; I39803

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Acevedo, V.; Ber
 C.: Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizel, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 V.: M.: Ogawa, K.; Ogilwara, C.; Rocha, E.; Roche, B.; Rose, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, S.H.; Parro, V.; Pohl, T.M.; Portete
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 aeneuhl, M.; Tamakoshi, A.; Tanaka, T.; Terpilte, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.: Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Yamamoto, H.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033
 A:Accession: 669593
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-508 <KUN>
 A:Cross-references: GB:299113; GB:AL009126; NID:92634090; PIDN:CA813696.1; PID:el1834

A:Experimental source: strain 168
 R:MacKay, R.M.; Lo, A.; Wlilick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; S
 Nucleic Acids Res. 14, 9159-9170, 1986
 A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
 A:Reference number: A26114; MUID:87066783
 A:Accession: A26114
 A:Molecule type: DNA
 A:Residues: 10-508 <MAC>
 A:Experimental source: strain PAP115
 A:Note: part of this sequence, including the amino end of the mature form, was confir
 R:Rindahl, V.; Aa, K.; Tronsmo, A.
 Antonio Van Leeuwenhoek 66, 327-332, 1994
 A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtit
 A:Reference number: I40353; MUID:95225656
 A:Accession: I40353
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-291, 'N', 293-508 <LIN1>
 A:Cross-references: EMBL:X67044; NID:939776; PIDN:CAA47429.1; PID:939777
 R:Rindahl, V.; Aa, K.
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S24239
 A:Accession: S24239
 A:Molecule type: DNA
 A:Residues: 10-291, 'N', 293-508 <LIN2>
 A:Cross-references: EMBL:X67044; NID:939776; PIDN:CAA47429.1; PID:939777
 A:Experimental source: strain CK-2
 R:Wolf, M.; Geczi, A.; Borris, R.
 submitted to the EMBL Data Library, December 1993
 A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtit
 A:Reference number: S49103
 A:Accession: S49103
 A:Molecule type: DNA
 A:Residues: 10-508 <NOL>
 A:Cross-references: EMBL:Z29076; NID:9509266; PIDN:CAA82317.1; PID:9509267
 R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:06:42; Search time 248.55 seconds
(without alignments)
76.597 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGYVSGTLLYDANGNP.....TIVNGPYGLRSTRSLSTVFYF 300

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 182106 seqs, 63460219 residues
Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_65:++
1: PIR1:++
2: PIR2:++
3: PIR3:++
4: PIR4:++

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1169	71.7	516	2 JEO134	mannan endo-1,4-beta
2	710	43.6	384	2 T37232	secreted beta-mann
3	692	42.5	363	2 S30386	mannan endo-1,4-beta
4	601	36.9	1331	2 A48954	mannan endo-1,4-beta
5	202	12.4	409	2 B25156	cellulase (EC 3.2.
6	200	12.3	488	2 A25156	cellulase (EC 3.2.
7	197	12.1	557	2 JCS487	cellulase (EC 3.2.
8	184.5	11.3	505	2 JCS962	endo-glucanase - Er
9	180	11.0	448	2 A27631	cellulase (EC 3.2.
10	176.5	10.8	466	2 C42360	cellulase (EC 3.2.
11	174.5	10.7	499	2 A27198	cellulase (EC 3.2.
12	174.5	10.7	504	2 A54744	cellulase (EC 3.2.
13	173.5	10.6	508	2 A26874	cellulase (EC 3.2.
14	162.5	10.0	486	2 I40548	bi-functional cellu
15	162.5	10.0	508	2 G65933	cellulase (EC 3.2.
16	157	9.6	783	2 JCS467	cellulase (EC 3.2.
17	152	9.3	800	2 A29003	cellulase (EC 3.2.
18	152	9.3	822	2 JTO611	cellulase (EC 3.2.
19	151.5	9.3	499	2 JTO111	cellulase (EC 3.2.
20	151.5	9.3	825	2 JSD174	cellulase (EC 3.2.
21	149.5	9.2	357	2 PC4404	cellulase (EC 3.2.
22	149.5	9.2	491	2 S28043	cellulase (EC 3.2.
23	147.5	9.0	429	2 S29044	endo-glucanase A pr
24	136.5	8.4	570	2 S56132	cellulase (EC 3.2.
25	131	8.0	814	1 C2C1EM	cellulase (EC 3.2.
26	130	8.0	426	2 A42649	cellulase (EC 3.2.
27	137	7.8	32	2 PC4278	guar gum-degrading
28	122.5	7.5	915	2 A48802	cellulase (EC 3.2.
29	122.5	7.5	1039	2 S02711	cellulase (EC 3.2.

30	122	7.5	438	2 A47702	glucan 1,3-beta-gl
31	117.5	7.2	428	2 S03767	cellulase (EC 3.2.
32	116	7.1	611	2 JCI177	endo-glucanase V (E
33	115.5	7.1	747	2 B47093	cellulase (EC 3.2.
34	114	7.0	419	2 S72325	glucan 1,3-beta-gl
35	113.5	6.8	517	2 I40798	cellulase (EC 3.2.
36	111.5	6.8	441	2 A44815	cellulase (EC 3.2.
37	111	6.8	475	1 C2C1CA	cellulase (EC 3.2.
38	110	6.7	26	2 PC4285	guar gum-degrading
39	109.5	6.7	754	2 T14877	hypothetical prote
40	107	6.6	430	2 S53325	endo-beta-1,6-glic
41	105	6.4	364	2 S12017	endo-glucanase A -
42	105	6.4	456	2 T40276	probable gas1 fam1
43	104	6.4	566	2 A40589	cellulase (EC 3.2.
44	101.5	6.2	317	2 B72216	endo-glucanase - Th
45	100.5	6.2	748	2 S19652	cellodextrinase C

ALIGNMENTS

RESULT 1
JEO134
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N: Alternate names: endo-1,4-beta-mannanase
C: Species: Bacillus circulans
C: Date: 03-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C: Accession: JEO134
R: Iosnida, S.; Sako, Y.; Uchida, A.
Biosci. Biotechnol. Biochem. 62, 514-520, 1998
A: Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A: Reference number: JEO134; MUID:98233274
A: Accession: JEO134
A: Molecule type: mRNA
A: Residues: 1-516 <POS>
A: Cross-References: DDBJ:AB007123
A: Note: the authors translated the codon CAA for residue 259 as Asn, CAA for residue
C: Keywords: glycoprotein; glycosidase; hydrolase

Query Match	71.7%	Score 1169;	DB 2;	Length 516;
Best Local Similarity	69.2%	Pred. No. 1e-83;		
Matches 207;	Conservative 48;	Mismatches 44;	Indels 0;	Gaps 0;
QY	1	NANSGYVSGTLLYDANGNP	PYVRGNGINCHANYKQOATTAEGIANTCANTYRIVLSDG	60
DB	33	HAASGYVSGTLLDLPATG	PYVRGNGINCHANYKQOATTAEGIANTCANTYRIVLSDG	92
QY	61	QWTKDHIHVRMLISAEEDNHLVAVPEVDATGYSISLNRVAVYTEMRSALIGKEDT	120	
DB	93	KMTLDVNTVNNILITCEQNKILAVLEVDATGSSLSLDNAVMTWIGIKSALIGKEDR	152	
QY	121	VIIINANEMWEGSMEDGAWADGYKQAIPLRNAGLNTHTLVYDAAGGQFOSIHDYGREVF	180	
DB	153	VIIINANEMWGTWDGVAWANGYKQAIPLRNAGLNTHTLVYDAAGGQFOSIHDYGREVF	212	
QY	181	NADPQNTMFSTHMYEYAGNAGSOVRNTINDRYLNDLALVIEFSEHRRNTGVDDEATMS	240	
DB	213	NADPQNTMFSTHMYEYAGNAGSOVRNTINDRYLNDLALVIEFSEHRRNTGVDDEATMS	272	
QY	241	YSEORGVGWLAWSKNGPEMEYLDLSNDWAGNNTLVAGNTIVNGPYGLRSTRSTVF	299	
DB	273	YSEORGVGWLAWSKNGSSDLAYLDXTNDWAGNNTLVAGNTIVNGPYGLRSTRSTVF	331	
RESULT 2				
T37232		secreted beta-mannosidase - Streptomyces coelicolor		
C: Species:		Streptomyces coelicolor		
C: Date:		03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999		
C: Accession:		T37232		
R: Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.				

This Page Blank (uspto)

Query Match Similarity 24.5%; Score 638; DB 2; Length 1331;
Best Local Smilarity 31.4%; Pred. No. 1,4e-35;
Matches 166; Conservative 84; Mismatches 208; Indels 70; Gaps 14;

QY 6 KRIFSTISLLASSILFVS-----GSTANANGSFVYSTGLTYDANGNPFVMRGIN 57
|| : :: : ||| : : : | : | :
DB 9 KKMLSVLCTIVPELNLFLFIANTVILPRVGATSN-----DVVKIDTS----TLIGTN 57
OY 58 HGHAWTKDOATAIEGIANTGANTVRIVLSDGGOWTKDDITFYANKLISLADNHLYA-VP 116
- - - - - : : : : : - - - - - : : : : : - - - - - : : : : :
DB 58 HAHCHWRDLRDLTALRGRIRSGMNSVRYVLSNGYRWTKIPASEVANIIISLSLGFKATIL 117
OY 117 EYHDATGY---DSIASUNRAVDYWIEMRSALIGKEDPVYIINIAEMFGSWEGDAMADCYK 173
|| : : : : : || : : : : : || : : : : : || : : : : :
DB 118 EYHDTTGDEGDACSLAQAVEYWKELKSVDLGDNEDEVIIINGNEPYNONNVNQWNTFK 177
OY 174 QAIPLRLNAGLNTLTAVDAAGMQ-FPOSIDHYGREVFADPNORTMFSIHKETAYAGNA 232
|| : : : : : || : : : : : || : : : : : || : : : : :
DB 178 NIKALRDAGFEKTIVDAENMGDMNSTMRDNAOSIMEADPLRLTVPSIHAYG-VYNTA 236
OY 233 SOVRTIIDRLNDLALVIEFGHRHTNGDVDEATINSYSORGTGMALMSKNGPEWE 292
- - - - - : : : : : ||| : : : : || : : : : : || : : : : :
DB 237 SKKEEYIKSFVDGLPLVIYEFGHQTDGDPDEBALVYKAQYKIGLEFSWSKGSSSYVG 296
OY 293 YLDLSDMAGNNLTAMGNTLVNGPYGLRETSRLSTVFTEGGSDDGT-SPTLLYDEGSNQ 351
|| : : : : : || : : : : : || : : : : : || : : : : :
DB 297 YLDVNMNMWDNNPTPMGWQRTNALIGTSPTPTSTVPTPLPTPTPTVATPTPTPT 356
OY 352 GMTGSSLG-----GPAVTEMSSGSHSKRADIQLSNSQHLYHI 393
- - - - - : : : : : - - - - - : : : : : - - - - - : : : : :
DB 357 PVSTPATSGCIKVLYANKETNSTNTIRPWL-----RVNSSGSSIDLSTRYTIKWTV 410
OY 394 QMTSLQONSIOATVFKANMGSVGNMTARLYYKIGH-----GYTWSGSEVPPI 442
DB 411 -----DGEPAQSRI--SDMAQIGASNTFRKYKLSSSVSGADYYLEIGFKSAGAGLQG 462
OY 443 NGSSGTTLSLDLSNVQNLSQVREIG-VQFSASDSSGGTSTIIDNYI 489
DB 463 KDTGEIOMRFNKMDMSMYNOGDMWSMIOSMTSYGENEKRYATYIDVLY 510

RESULT 5
A:J219
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus sp. (strain AM-001)
N:Alternate names: 1,4-beta-d-mannan mannohydrolase; beta-mannanase
C:Species: Bacillus sp.
C>Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 15-Oct-1999
C:Accession: A37219; B37219; P10111
R:Amino, T.: Kato, C.; Horikoshi, K.
A>Title: Two Bacillus beta-mannanases having different COOH termini are produced in Esch
A:Reference number: A37219; MUID:90146329
A:Accession: A37219
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1513 <AKI>
A:Cross-references: GB:M31797; NID:g143166; PIDN:AAA22586.1; PID:g143167
A:Accession: B37219
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <AK2>
A:Cross-references: GB:M31797
R:Amino, T.: Kato, C.; Horikoshi, K.
Arch. Microbiol. 152, 10-15, 1989
A>Title: The cloned beta-mannanase gene from alkalophilic Bacillus sp. AM-001 produces t
A:Reference number: P10111
A:Accession: P10111
A:Molecule type: protein
A:Residues: 30-41 <AK3>
A>Note: three forms of the enzyme (M-I, M-II, and M-III) with different molecular weight
A>Note: all three forms of the enzyme have the identical amino-terminal sequence shown h

C:/function: cleaves the beta-1,4-mannosidic linkages in various beta-mannans
A:description: glycosylase; hydrolase; polysaccharide degradation
C:keywords: glycosylase; hydrolase; polysaccharide degradation

Query Match 16.48; Score 427; DB 2; Length 513;
Best Local Similarity 26.9%; Pred. No. 9.6e-22;
Matches 149; Conservative 76; Mismatches 215; Indels 114; Gaps 23;

QY 5 EKRTSTLSLLASSILFVSGTSTANANSFPYSGTLYANGNPFVM----- 53
D 4 YKKVFVAIAFFEFVLPTLTISSSEAN-----GAALLSPNANOTTKNYSMLALPN 55
QY 54 ---RGINHG-AMYDKOATTAEGIANTGANTVRIVLS-DGGOW----TKDDIHVFN- 102
D 56 KSNKRVYSGHFGCYSDSLAWIKOCARELTGMPEILSCDKNNQTRLYVADQLSYGCNQ 115
QY 103 -LISTAEADNHAAV-----PEVDATGYDSIASLNRAVDY-----WIEKRSALI 145
D 116 ELINFMNOGLVTISVHPNPGFHSGENYKTLLPFSOFONLTNHTTEGRMKMDLMXA 175
QY 146 GNED-----TVLINANETGSMEGDAW--ADGKQAIRLRAGLN-----H 185
D 176 DGIDELONNGVTVLEFRPLKENGEM--FWMGAEQYNPDFOTRANAATISMRDMYQIFTH 232
QY 187 TLMVPAAGGCGPPSIHDGREVFANDPORNTMFSIIHYEYAGNASOVRTINDREVNOD 246
D 233 EKKLNKLTVYSPDYRRH---VTSYYPKANVVDIALDSY-HRPRLSLTDQYNNMILD 288
QY 247 LALVIEFGHRHT-NGDVDEATIMSYSEQ--RGVGWLAMS--WK--GNGPWEYLDSL 298
D 289 KPFAAFELIGPPSMAGSFYSNYIAIKOKYPRTYFLAMNDKMSPHNNRCAG--DLFN 345
QY 299 DNAGNNLTAMGNTIYNGPYGLAETSRLSTVPFGGSDGTS--PTTVIDEESMGWTGS 356
D 346 D-----SWVNRGEIDYGQSPATVLLDFENNLISMGC 379
QY 357 SLS-GGPMVAITEMSSKSGSHLKADIOLSSNSOHYLVHIONTLQOONSRIQATVKHANMS 415
D 380 EPFDGGPMVSNEKSANGTOSLKAADVILGNNSYH-LQKYNRRLSSFKNELIVSHSSMN 438
QY 416 VGNGTARLYVKTGHGYTWYSGSEVPINGSSGTTSLDLSNVONTLSQVEICVQFSAS 475
D 439 VGSGMARAFVFKKGSAWRNNAGEFCOFAGRRTALTSLDTKYSNLHDVRELIVEYKAPAN 498
QY 476 SSGQTSTIIDNVIV 489
D 499 SNGKTAIYLDHYTV 512

RESULT 6
JCS487
cellulase (EC 3.2.1.4) precursor - Clostridium cellulovorans
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium cellulovorans
C:date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 15-Oct-1999
R:Stewelt, S.A.; Ichii-Ishi, A.; Park, J.S.; Liu, C.; Malburg, L.M.; Dol, R.H.
Gene 182, 163-167, 1996
A:title: Characterization of engF, a gene for a non-cellulosomal Clostridium cellulovorans
A:reference number: JCS487; NCBI:57136706
A:accession: JCS487
A:molecule type: DNA
A:Residues: 1-557 <HEH>
A:Cross-references: GB:U37056; NID:g1778708; PID:NAB40891.1; PID:g1778709
A:Experimental source: strain ATCC 35296
A:Accession: PC4333
A:molecule type: protein
A:Residues: 30-37 <SHEZ>
C:Comment: This enzyme plays a secondary role in cellulose degradation.
C:Genetics:
C:Keywords: glycosylase; hydrolase; polysaccharide degradation

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:32:55 ; Search time 362.48 Seconds
(without alignments)
77.279 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGEYVSGTLLYDANGNP.....TIVNGPYGLRENSRLSTVPT 300

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues
Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_14:*
2: SP_Archaea:*
3: SP_Bacteria:*
4: SP_Fungi:*
5: SP_Human:*
6: SP_Invertebrate:*
7: SP_Mammal:*
8: SP_Protista:*
9: SP_Phage:*
10: SP_Plant:*
11: SP_Rodent:*
12: SP_Virus:*
13: SP_Vertebrate:*
14: SP_Unclassified:*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1172	71.9	516	2 066185	066185 bacillus ci
2	797	48.9	395	2 069347	069347 vibrio sp.
3	710	43.6	384	2 086599	086599 streptomyc
4	634	38.9	279	2 092F13	092F13 thermomonos
5	614	37.7	2	09RFX5	09RFX5 calidbacill
6	536.5	32.9	327	2 09RJR5	09RJR5 streptomyc
7	226	13.9	1097	2 09ZAI7	09ZAI7 anaerobic t
8	197	12.1	557	2 P94622	P94622 clostridium
9	191	11.7	389	2 059232	059232 bacillus sp
10	176.5	10.8	501	2 083012	083012 bacillus su
11	174.5	10.5	499	2 045532	045532 bacillus su
12	170.5	10.2	387	2 059290	059290 clostridium
13	166.5	10.0	387	2 031029	031029 erwinia car
14	162.5	9.9	486	2 045430	045430 bacillus sp
15	161.5	9.9	499	2 052731	052731 bacillus sp
16	160.5	9.8	481	2 066064	066064 actinomyc
17	159.5	9.8	635	2 066065	066065 fibrobacter
18	157	9.6	783	2 045554	045554 bacillus sp
19	154.5	9.5	749	2 059154	059154 anaerocellu

20	152	9.3	821	2 059241	059241 bacillus sp
21	140	8.6	278	2 09RKS5	09RKS5 streptomyc
22	136.5	8.4	570	2 059665	059665 pseudomonas
23	129	7.9	478	5 016028	016028 globodera r
24	126	7.7	319	5 018454	018454 heterodera
25	126	7.7	438	3 09URL8	09URL8 candida alb
26	126	7.7	476	5 018453	018453 heterodera
27	125	7.7	395	5 0906M4	0906M4 globodera t
28	123	7.5	319	5 061595	061595 heterodera
29	123	7.5	319	5 077449	077449 heterodera
30	121.5	7.5	1000	2 024820	024820 thermophil
31	117	7.2	391	5 077094	077094 globodera r
32	117	7.2	392	5 044078	044078 globodera r
33	116.5	7.1	332	2 060054	060054 undentifite
34	116	7.1	426	2 09REW0	09REW0 erwinia chr
35	115	7.1	363	2 007652	007652 celivibrio
36	114	7.0	419	3 012539	012539 agaricus bl
37	114	7.0	470	5 0906M5	0906M5 globodera t
38	113	6.9	910	3 087211	087211 orpionmyes
39	112.5	6.9	621	2 007653	007653 celivibrio
40	111.5	6.8	506	5 090A57	090A57 meloidogyne
41	110	6.7	435	3 012626	012626 pichia angu
42	109.5	6.7	754	2 085318	085318 salmonella
43	107	6.6	430	3 012712	012712 trichoderma
44	106	6.5	325	3 09Y8H6	09Y8H6 emericella
45	105	6.4	456	3 09Y7Y7	09Y7Y7 schizosacch

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	516 AA.
066185	066185			
AC	066185;			
DT	01-AUG-1998 (TRENBLREL. 07, Created)			
DT	01-AUG-1998 (TRENBLREL. 07, Last sequence update)			
DT	01-MAY-2000 (TRENBLREL. 13, Last annotation update)			
DE	MANNANASE.			
OS	Bacillus circulans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
CC	Bacillus/Staphylococcus group; Bacillus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98233274.			
RA	Yoshida S., Sako Y., Uchida A.;			
RT	Cloning, sequence analysis, and expression in Escherichia coli of a			
RT	gene coding for an enzyme from Bacillus circulans K-1 that degrades			
RT	guar gum."			
RL	Biosci. Biotechnol. Biochem. 62:514-520(1998).			
DR	EMBL; AB007123; BAA25878.1; -			
DR	INTERPRO; IPR001547; -			
DR	PFAM; PF00150; cellulase; 1.			
SO	SEQUENCE 516 AA; 55245 MW; DBE5A48842AE8062 CRC64;			
Query Match	71.9%; Score 1172; DB 2; Length 516;			
Best Local Similarity	69.2%; Pred. No. 2e-84;			
Matches	207; Conservative 50; Mismatches 42; Indels 0; Gaps 0;			
QY	1 NANSGEYVSGTLLYDANGNPVFGNGINHGAMVYKDQATTAEGIANGTAVRIYLSGG 60			059241 bacillus sp
DB	33 HAASGEYVSGTLLYDANGNPVFGNGINHGAMVYKDQATTAEGIANGTAVRIYLSGG 92			09RKS5 streptomyc
QY	61 QMTDDITFVNLISLADNHLVAPEVHADTGYDSLSRAVDYWTSEMSALIGKEDT 120			059665 pseudomonas
DB	93 KMTLDVATVNNITLCEQNKILAVLEVHADTGSDSLSDNNAYVWIGISALIGKEDR 152			016028 globodera r
QY	121 VIINANEMFGSWGDADWDYKQAIPLRNAGINHTLMDVDAAGGQFPOSIDHYGREVF 180			018454 heterodera
DB	153 VIINANEMFGSWGDADWDYKQAIPLRNAGINHTLMDVDAAGGQFPOSIDHYGREVF 212			09URL8 candida alb
QY	181 NADPQNTMESIHETVYAGNAGNAGVNTINIDRYLNMOLALVIGFGHRTNGDVDEATIMS 240			018453 heterodera

DB 213 NADPLKNTFVSIHMEYTAGNASTYKSNIDCVLNKLNLLIIGERGAGGTNDVDDEATLMS 272
QY 241 YSEORGVCWGLAMSWKGNPEWEYDLSDNAGNNLTAMGNTIYNGPYGLRSTLSTVE 299
DB 273 YSQERGVCWGLAMSWKGNSSDLATLDMTNDMAGNSLTSGNTVNGSNKIKATSVLSGIT 331

RESULT 2
ID 069347 PRELIMINARY: PRT: 395 AA.
AC 069347
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
GN BETA-1,4-MANNANASE.
OS Vibriol sp.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibri-
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MA-138;
RA Tamaru Y., Arai T., Morishita T., Kimura T., Sakka K., Ohmura K.;
RL J. Ferment. Bioceng. 83:201-205(1997).
DR EMBL; D86329; BAA25188.1; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002883; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF02013; CBD.5; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 395 AA; 43098 MW; A3E30B3A978C5DA1 CRC64;

Query Match 48.9%; Score 797; DB 2; Length 395;
Best Local Similarity 50.8%; Pred. No. 4.4e-55;
Matches 155; Conservative 55; Mismatches 87; Indels 8; Gaps 5;

QY 2 ANSGFYVSGTLLYDANGNPFVNRGINHGHWKQDATTALIEGANTGANTVRYLSGCG 61
DB 19 AAGGIVHSNGVLYEANGSAFKIRGINHATWYTDKLSVALSGIATGANTVRYVLSNGR 78
QY 62 WTKDHIYVRNLISLAEDNHLVAPEVHDATGY--DSIASNRAVDYIEMRSALIGKE 118
DB 79 WTKNVSDVTNLTINAKANNLTALIEVHDTCYGERSSAASLDSADTYIELKNEIIGE 138
QY 119 DTVIINIANEMFG-SWEGDAMADGYKQALPRLNAGNLNHTLVNDAAGWGQ-EPGSHDYG 176
DB 139 DVIYINLGNPEFGNNNDAAVAVNDHVSATQRLRSAGINHTIWDAPNMGQDWKGFMLNA 198
QY 177 REVENADPQRNTMESIHMEYTAGNAGYRINIDRYLNQDLALVYGERGHRITNDVDE 235
DB 199 QPVEFMSDPLNITFVHMEYVSYNS--VNDYISFTNGVLVYGERASTHKADVDE 256
QY 236 ATIMSSEORGVCWGLAMSWKGNPEWEYDLSDNAGNNLTAMGNTIYNGPYGLRSTL 295
DB 257 GSIMERSETLSGTYIGMSNGSDTTSDLDIVANNMNNISYTWGNVLINGNGISTSTL 316
QY 296 STVEFT 300
DB 317 ATVEFT 321

RESULT 3
ID 086599 PRELIMINARY: PRT: 384 AA.
AC 086599
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
GN SECRETED BETA-MANNOSIDASE.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RT Kinash H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031514; CAA20610.1; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002883; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF02013; CBD.5; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 384 AA; 39847 MW; B8AD1583FD3BD11 CRC64;

Query Match 43.6%; Score 710; DB 2; Length 384;
Best Local Similarity 46.7%; Pred. No. 3e-48;
Matches 141; Conservative 53; Mismatches 102; Indels 6; Gaps 4;

QY 2 ANSGFYVSGTLLYDANGNPFVNRGINHGHWKQDATTALIEGANTGANTVRYLSGCG 61
DB 37 AAGGIVHSNGVLYEANGSAFKIRGINHATWYTDKLSVALSGIATGANTVRYVLSNGR 95
QY 62 WTKDHIYVRNLISLAEDNHLVAPEVHDATGY--DSIASNRAVDYIEMRSALIGKE 118
DB 96 WTKNSASEVSNALIOCKANKVYCLEVHDTCYGERGASLSDADTYVSYKALSIGE 155
QY 119 DTVIINIANEMFG-SWEGDAMADGYKQALPRLNAGNLNHTLVNDAAGWGQ-EPGSHDYG 177
DB 156 DVIYVINGNEFGNTATAMDATKSAIGLRAGGLDHALVNDAPNMGQDWKGFMSNA 215
QY 178 EVENADPQRNTMESIHMEYTAGNAGYRINIDRYLNQDLALVYGERGHRITNDVDEAT 237
DB 216 VEFASDPEPRNTVEFVHMG-VYDTAAEVRDYLNAFVSGLEPIYVGERGSDHSDENPEDA 274
QY 238 IMSSSEORGVCWGLAMSWKGNPEWEYDLSDNAGNNLTAMGNTIYNGPYGLRSTL 297
DB 275 IMATROSLGVLYLMSNGSGNGGVEYLDVANGFDPNSLTSGNLTIFGSNGIATSTRTAT 334
QY 298 VE 299
DB 335 VY 336

RESULT 4
ID 092F13 PRELIMINARY: PRT: 279 AA.
AC 092F13
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
GN BETA-MANNANASE (EC 3.2.1.78) (FRAGMENT).
OS Thermomonaspora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KW3;


```

Db 215 STANGETEONTPTVDKEKRSAYLTYYGDDLNADBSAKRTTAFENYN-EGDIQKVEDYIDR 273
Oy 212 VIANODLALVIGEGHRTTNGDVDEATIMSYSE--ORGYGLWASH-----KGN 257
Db 274 ANMAGLVFVEEYKGDYS--DAAEGKSGSLQVNMKKGAGRIYMWMDGYDLTISGTR 331
Oy 258 GPWEYIDLSNDMAGNNLTAMGNTIYNGPYGLAET 292
Db 332 GSGWE-INKTDGSKPTNLWSWGDKINDNNGIIFT 365

RESULT 8
P94622 PRELIMINARY; PRT; 557 AA.
AC P94622.
DT 01-MAY-1997 (TIMBLREL 03, Created)
DT 01-MAY-1997 (TIMBLREL 03, Last sequence update)
DT 01-MAY-2000 (TIMBLREL 13, Last annotation update)
DE ENDO-1,4-BETA GLUCANASE ENGf (EC 3.2.1.4) (CELLULASE) (ENDOGALACTANASE; (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE)).
GN ENGf.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
OC [1]
RN RN
RP SEQUENCE FROM N.A.
RX MEDLIN: 97136706.
RA Shevela S.A., Ichl-Ishl A., Park J.S., Liu C., Malburg L.M.,
RA Dol R.H.;
RA "Characterization of engf, a gene for a non-cellulosomal Clostridium cellulovorans endoglucanase.";
RL Gene 167:163-167(1996).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.
CC EMBL: U37056; AAB40891.1; -.
DR HSSP: 085465; ZASH.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Hydroxylase; Glycosidase.
SQ SEQUENCE 557 AA; 60131 MW; D186EC88EB504EED CRC64;

```

```

RESULT 9
059232 ID 059232 PRELIMINARY; PRT; 389 AA.
AC 059232;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 13, last annotation update)
DE ENDO-BETA-1,4-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULOSE)
DE ENDO-BETA-1,4-GLUCANASE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-186-1;
RA Sanchez Torres J.;
RL Thesis (1994), Universidad de Salamanca, Spain.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
DR EMBL; Z3876; CA863942.1; -.
DR HSSP; 085465; 2A3H.
DR INTERPRO; IPRO01547; -.
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 389 ENDO-1,4-BETA-GLUCANASE.
SQ SEQUENCE 389 AA; 43688 MW; 9111397485609B5F CRC64;

```

```

Query Match 11.7%; Score 191; DB 2; Length 389;
Best Local Similarity 25.1%; Pred. No. 2.2e-07;
Matches 70; Conservative 45; Mismatches 102; Indels 62; Gaps 13;

```

```

QY 8 VSGTLYDANGNPFVARGI-NHGAWYKQDATTAEIGANTGANTVRIYVSDGGQTKDD 66
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 40 IINGELVNRGRQVQLKGSISHLQWTG-----QFNYESMKLRDD 81
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 -----IHT-----VRNLISLAEDNHLVAPEVDATGYDSIASLNPRA 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 82 WGITVFRAMYTSSGGYIDDPVKEKVEAVEAIDLDIYIIDHILSDNDPNYKEEA 141
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 104 VQVYIEMRSALIGKEDYVIINIANEFSGWEGDANADGRK---QAIPLRNAGLNHTLM 159
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 142 KQFEDEM-SELYGDPNVPYIEIANEPNGS--DYTWNRKRPAAEYIPYIRNNDPNNII 198
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 160 VDAAGMGQPOSIDHYGREVFANADPORNTFMSIHMYEYAGNASOVRTIDRYLNDLAL 219
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 199 V---GTGTWSQVH-HAADNQLADP--NMYIAFHF--YAGTGHQNLDDQVYALDQCAAI 250
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 VIGFEGHRTNGD---VDEA-TIMSYSEORGVGLAWS 253
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 251 FVSEWGTSAATGQGVFLDEAQEWIDFMDERNLSTANWS 289
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 10
083012 ID 083012 PRELIMINARY; PRT; 501 AA.
AC 083012;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
DE CELLULOSE.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-5H;
RA Khamonuch C., Ooi T., Kinoshta S.;
RT "Cloning and nucleotide sequence of beta-mannanase and cellulase gene
   from Bacillus sp. 5H.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBD databases.

```

```

DR EMBL; AB016164; BA31712.1; -.
DR HSSP; 085465; 2A3H.
DR INTERPRO; IPRO01547; -.
DR INTERPRO; IPRO01956; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00942; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PRODOM; PD001947; -. 1.
SQ SEQUENCE 501 AA; 55422 MW; F76D260A901E2D1D CRC64;

```

```

Query Match 10.8%; Score 176.5; DB 2; Length 501;
Best Local Similarity 21.1%; Pred. No. 4.4e-06;
Matches 70; Conservative 67; Mismatches 122; Indels 73; Gaps 15;

```

```

QY 3 NSGFVSGTLYDANGNPFVARGI-NHGAWYKQDATTAEIGANTGANTVRIYVSDGGQ 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 39 NGSLSTKGTQLNQNKANQVLKGISISHLQWTG-----FVNKDSLK 80
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 WTRKD-----IHT-----VRNLISLAEDNHLVAPEVDATGYDSIA 98
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 81 WLRDDGITVFRAMYTADGGYIDNPSRKRYKAVEAKELGIYIIDHILNDGNPNQ 140
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 99 SLNRADVYIEMRSALIGKEDYVIINIANEFSGWEGDANADGRK---QAIPLRNAGLN 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 141 NKEKAFFKEMF-LYGTWPVYIEIENEPNGVDWKKRDIPTVAE---EVISVIRKNDP 196
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 NHTLWDAAGMGQPOSIDHYGREVFANADPO---RNTMFSIHMYEYAGNASOVRTIDR 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 197 DNTIIV---GTGTWSQVH-----AADQLKDNVYIALHF--YAGTGHGSISDKANY 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 VINQDLALVIGFEGHRTNGD---VDEA-TIMSYSEORGVGLAWSKNGPWEYL-- 264
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 246 ALSKGAPIFVTEWGTSDASNGGVFLDQGREMLANTYDSKRISVMVNLSDKQESSALKP 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 265 --DLSNDAGNNLTAMGNTIVAGPYGLRSTR 294
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 306 GASKTGWPLTDLTASGTFVRENIFGNKDKSTR 337
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 11
045532 ID 045532 PRELIMINARY; PRT; 499 AA.
AC 045532;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
DE CELLULOSE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87190397.
RA Nakamura A., Uozumi T., Teruhiko B.;
RT "Nucleotide sequence of a cellulase gene of Bacillus subtilis.";
RL Eur. J. Biochem. 164:317-320(1987).
DR EMBL; M28332; AAA22307.1; -.
DR HSSP; 085465; 2A3H.
DR INTERPRO; IPRO01547; -.
DR INTERPRO; IPRO01956; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00942; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PRODOM; PD001947; -. 1.
SQ SEQUENCE 499 AA; 55075 MW; D424AB9E63B94D23 CRC64;

```

```

Query Match 10.7%; Score 174.5; DB 2; Length 499;
Best Local Similarity 21.1%; Pred. No. 6.2e-06;
Matches 70; Conservative 66; Mismatches 123; Indels 73; Gaps 15;

```

```

QY 3 NSGFVSGTLYDANGNPFVARGI-NHGAWYKQDATTAEIGANTGANTVRIYVSDGGQ 61

```

```

Db      39 NQSLIKGTQVLRDGAQVQKLGISSHGLQWYGD-----FYNKDSLK 80
QY      62 WKKD-----IHT-----VRNLISLAEDNHLVAVPEVDATGYDLSIA 98
Db      81 WLRDWMGITVFRAMATTADGGYIDNPVSKNKYKEAVEAKELGIYVLDHMLDGNPNQ 140
QY      99 SLNRVADYWIEMRSALIGKEDTVIINIANEMFG--SWEGD--AMADGYKQALPRLRNGL 154
Db      141 HHEKADPFKEM--SLYGNTPNVIYEIANEPGVDNWKMDIRPYA---EYISYIRKNDP 196
QY      155 NRTIAMDAGNGQFPOSIIIDYGREVFNADPO---RNTMFSIMYEYAGGNASOVRTNIDR 211
Db      197 DNTIIV---GTGTWSDVDND-----AADQDKANVYALHF--YAGTHGSLDKXANY 245
QY      212 VLNODLALVIGERGHHTNGD---VDEA--TMSYSEORGWGLMSWKNGPEWEYL-- 264
Db      246 ALSKGAIPITVETENGSTDASNGGVFLDQSRWMLNTYDSKNISWYMNLSLDRKQESSALKP 305
QY      265 --DLSDMAGNNTIANGNTIVNGPYGLRETSR 294
Db      306 GASKTGWPLTDLTASGTFRENILANKDSTK 337

```

```

RESULT 12
Q59290 PRELIMINARY; PRT; 930 AA.
AC Q59290;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, last annotation update)
DE BETA-1,4-ENDOGLUCANASE (EC 3.2.1.4).
GN CELA.
OS Clostridium jostii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group: Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RA Fujino T., Fujino E., Karita S., Ohmura K.;
RT "Revised sequence of cels gene encoding endoglucanase (Eg)-1 from
RL Clostridium jostii";
RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; D85326; BA12826.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001119; -.
DR INTERPRO; IPR001547; -.
DR PRAM; PF00150; cellulase; 1.
DR PRAM; PF00395; SLH; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR Hydrolase; glycosidase.
KM SEQUENCE 930 AA; 101726 MW; 0E8AE8754D084448 CRC64;
SQ

```

```

Query Match 10.5%; Score 170.5; DB 2; Length 930;
Best Local Similarity 26.0%; Pred. No. 3.2e-05;
Matches 78; Conservative 45; Mismatches 106; Indels 71; Gaps 19;

```

```

QY      12 TLYDANGNPFVANGI--HGANY-----KQATTAEGTANGANTVRYVL--SSGGQWT 64
Db      61 TLCDKGNPQLRGSTHGLQWFPETINNAPALS--KDWSSNVIRLAMTVABSGYSKD 118
QY      65 DDIHYVRL--ISLAEDNHLVAVPEVDATGYDLSI--AVDYWIEMRSALIGKEDTV 121
Db      119 PELIKRIVDGLDIALANDVIYVDHVTLPDPPNADVYKGMDFKEL--SQKYNNPHI 177
QY      122 IINIANEMFGWEG--DA--WA--DGYKQALP--LRNAGLHTIAMDAGNGQFPOSIIH 173
Db      178 IYLANEPSPNDGVTNNDAGNAKVSYPEIKILIRDSGNKNTLIVGSPNNSQRP---- 233
QY      174 DYREVFNADPQNTMFSIMYEYAGGNASOVRTNIDRYLNDLALVIGEFHRTNGDV 233
Db      234 DLAAE--NPINDNNTAIVSFHFYSGT-----HKSTDSST 264

```

```

QY      234 DEATIMS---YSEORGWGLANSW-----KGNP-----EW-FYLDLSN--DMAGNLT 276
Db      265 DRGNIMSNARYLHGVAVFCEMGTSEASGNNGPYLKEADEWLEFELANNISWINSLSI 324

```

```

RESULT 13
Q31029 PRELIMINARY; PRT; 387 AA.
ID Q31029;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)
DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
GN CELA.
OS Erwilia carotovora subsp. carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwilia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT34;
RC Paik Y.W., Lim S.T., Yun H.D.;
RL Mol. Cells 0:0-0(1997).
DR EMBL; AF025768; AAC02964.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001547; -.
DR PRAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR Signal; Hydrolase.
KW CHAIN 1
FT SIGNAL 31
FT CHAIN 32
FT SEQUENCE 387 AA; 42003 MW; 6302B3DF8AC2738B CRC64;
SQ

```

```

Query Match 10.2%; Score 166.5; DB 2; Length 387;
Best Local Similarity 21.5%; Pred. No. 1.8e-05;
Matches 68; Conservative 50; Mismatches 113; Indels 85; Gaps 13;

```

```

QY      13 LYDANGNPFVANGI--NHGANYKQATTAEGTANGANTVRYVLSGGQWTKDD----- 66
Db      48 LVDEGKRVQLRGVSASHLQWFGD-----YRKSKSMWMLPDWGINV 89
QY      67 -----IHTVRNLISLAEDNHLVAVPEVDATGYDLSI--NRADVWI 108
Db      90 SRVAMTADGYISKPSLANKYKVAEAAQSLGYITIIDHMLSDNPNRIYKQAKTFFA 149
QY      109 EKRSALIGKEDTVIINIANEMFG--SWEGD--AMADGYKQALPRLRNAGLHTIAMDAG 164
Db      150 EM-AGLYGNSPIVIEIANEPGVTWDEIRPYA---LEVETIRSKDPNLTIV--G 202
QY      165 WQFPOSIIIDYGREVFNADPQNTMFSIMYEYAGGNASOVRTNIDRYLNDLALVIGEF 224
Db      203 TGTWSDIHDH---AADNQLPDNTLYALHF--YAGTHQFLRIRIDYQSGAALFVSEW 257
QY      225 GHRHTNGD---VDEATIMSSEORGWGLMSWKNGPEWEYLDLS----- 267
Db      258 GTSDASGNGPFLPESHHTYIDLINRG-----GSRVNSLSDKSEASALAPGASK 308
QY      268 --NDMAGNNTIANGNTI 282
Db      309 CGGWTBQNLASAGKFEV 324

```

```

RESULT 14
Q45430 PRELIMINARY; PRT; 486 AA.
ID Q45430;
AC Q45430;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
GN CEL.
OS Bacillus sp.

```



```

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D04;
RX MEDLINE; 96029707.
RA Han S.J., Yoo Y.J., Kang H.S.;
RT "Characterization of a bifunctional cellulase and its structural gene.
RT The cell gene of Bacillus sp. D04 has exo- and endoglucanase
RT activity."
RL J. Biol. Chem. 270:26012-26019(1995).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; U27084; AAC43478.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR001956; -.
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PRODOM; PD001947; -. 1.
RW Signal.
FT CHAIN 1 29 POTENTIAL.
FT DOMAIN 30 486 ENDOGLUCANASE.
FT ACT_SITE 169 486 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 257 486 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 486 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 486 AA; 53820 MM; BD3B65BF18D656B CRC64;

Query Match 10.08; Score 162.5; DB 2; Length 486;
Best Local Similarity 21.08; Pred. No. 5.3e-05;
Matches 66; Conservative 75; Mismatches 136; Indels 37; Gaps 15;

OY 3 NSGFVSGTLYDANGNFVARGI-NHGHWYKQAT-TATEGIANTGANTV---RIVLS 57
DB 39 NGQISIKTQLVNRDGAQVOLKGISSHGLWYGEIVNKSILKWLDMGIVTFRAMATTA 98
OY 58 DGGQWTKDID-HVYANLISLAEDNHLVAPEVHDATGYDSIASINRAVDYIEMKRSALIG 116
DB 99 DGGYIDNPVANKYKAEVAEAKELGIYIIDWHLIDGNPNQNEKAKEFEKEM-SSLYG 157
OY 117 KEDYIITIANEMWG--SWEGD--AMADGYKQALPRLNAGLINTLMDAAGWGFPOST 172
DB 158 NTPNVIYEIANEPNQDVNMKDKIPYAE--EVISIRKNDPDNIIV---GTGTWSQDV 211
OY 173 HDYGEVFNADPO--RRTMFSIHMYEYAGNASOVRNIDRYLNQDLALVIGEGHRT 229
DB 212 ND-----AADQQLKANAVMTALHF--YAGTHGQFLRKANVALSKGAPITVTEGTSDA 263
OY 230 NGD-----VDEA-TTMSYSEQGVGLTMSWKGNGPEWEYL-----DLSNDWAGNNLTANGN 280
DB 264 SGNGVFLDQSRWKLYDSTISWVNMNLSDKQESSSALKPGASKTGWRSLDLSASGT 323
OY 281 TTVNGPYGLRETSR 294
DB 324 FVRENILGTKDSTK 337

RESULT 15
ID 052731 PRELIMINARY; PRT; 499 AA.
AC 052731;
DT 01-JUN-1998 (TReMBLrel. 06; Created)
DT 01-JUN-1998 (TReMBLrel. 06; Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14; Last annotation update)
DE ENDO-B-1,4-GLUCANASE (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
GN CELS.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;

```

```

OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-79-23;
RA Jung K.H., Chun Y.C., Lee J.-C., Kim J.H., Yoon K.-H.;
RL Biotechnol. Lett. 18:1077-1082(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-79-23;
RA Yoon K.-H., Jung K.H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
DR EMBL; AF045482; AAC02536.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR001956; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00942; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PRODOM; PD001947; -. 1.
RW Hydrolase; Glycosidase.
SQ SEQUENCE 499 AA; 55193 MM; 990A1878788CF738 CRC64;

Query Match 9.98; Score 161.5; DB 2; Length 499;
Best Local Similarity 21.08; Pred. No. 6.6e-05;
Matches 66; Conservative 74; Mismatches 137; Indels 37; Gaps 15;

OY 3 NSGFVSGTLYDANGNFVARGI-NHGHWYKQAT-TATEGIANTGANTV---RIVLS 57
DB 39 NGQISIKTQLVNRDGAQVOLKGISSHGLWYGEIVNKSILKWLDMGIVTFRAMATTA 98
OY 58 DGGQWTKDID-HVYANLISLAEDNHLVAPEVHDATGYDSIASINRAVDYIEMKRSALIG 116
DB 99 DGGYIDNPVANKYKAEVAEAKELGIYIIDWHLIDGNPNQNEKAKEFEKEM-SSLYG 157
OY 117 KEDYIITIANEMWG--SWEGD--AMADGYKQALPRLNAGLINTLMDAAGWGFPOST 172
DB 158 NTPNVIYEIANEPNQDVNMKDKIPYAE--EVISIRKNDPDNIIV---GTGTWSQDV 211
OY 173 HDYGEVFNADPO--RRTMFSIHMYEYAGNASOVRNIDRYLNQDLALVIGEGHRT 229
DB 212 ND-----AADQQLKANAVMTALHF--YAGTHGQFLRKANVALSKGAPITVTEGTSDA 263
OY 230 NGD-----VDEA-TTMSYSEQGVGLTMSWKGNGPEWEYL-----DLSNDWAGNNLTANGN 280
DB 264 SGNGVFLDQSRWKLYDSTISWVNMNLSDKQESSSALKPGASKTGWRSLDLSASGT 323
OY 281 TTVNGPYGLRETSR 294
DB 324 FVRENILGTKDSTK 337

```

Search completed: December 19, 2000, 17:32:58
 Job time: 1525 sec

Wed Dec 20 15:36:52 2000

us-09-339-159-2_copy_31_330.rpt

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:04:10 ; Search time 226.76 seconds

(without alignments)
45.238 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630

Sequence: 1 NANSGRFYSGTLYDANGNP.....TIVNGPYGLRETSRLSTVPT 300

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

268485

Total number of hits satisfying chosen parameters:

268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.36:*

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1630	100.0	490	21	Y54122	Amino acid sequenc
2	1624	99.6	476	21	Y54123	A mannase-linker
3	1283	78.7	468	21	Y44496	Bacillus agaradher
4	1283	78.7	468	21	Y54125	Amino acid sequenc
5	1283	78.7	493	21	Y44495	Bacillus agaradher
6	1283	78.7	493	21	Y54124	Amino acid sequenc
7	1136	69.7	331	21	Y54127	Amino acid sequenc
8	1075	66.0	369	21	Y54129	Amino acid sequenc
9	937.5	57.5	320	21	Y54132	Amino acid sequenc
10	522	32.0	188	21	Y54133	Amino acid sequenc
11	369	22.6	305	21	Y54130	Amino acid sequenc
12	318	19.5	132	21	Y54131	Amino acid sequenc

13	202	12.4	410	18	W12378
14	202	12.4	411	18	W12379
15	202	12.4	411	18	W12381
16	202	12.4	412	18	W12380
17	201	12.3	409	18	W23601
18	201	12.3	400	19	W22521
19	201	12.3	400	19	W57431
20	201	12.3	462	19	W57433
21	200	12.3	409	9	P81843
22	191	11.7	467	17	W05731
23	191	11.7	467	17	W00382
24	174.5	10.7	499	14	R42122
25	171	10.5	352	20	Y08472
26	156	9.6	1010	19	W34989
27	152	9.3	800	8	P70420
28	152	9.3	822	13	R26021
29	149.5	9.2	357	16	R77394
30	149.5	9.2	941	11	R07478
31	149.5	9.2	941	16	R77395
32	142	8.7	358	19	W39262
33	142	8.7	521	17	R89927
34	142	8.7	562	21	Y69508
35	135	8.3	551	18	W18790
36	127	7.8	472	19	W37243
37	126	7.7	302	19	W43909
38	124	7.6	484	19	W43910
39	118.5	7.3	956	19	W48874
40	118	7.2	476	19	W37241
41	112.5	6.9	360	19	W49870
42	112.5	6.9	360	19	W37242
43	111.5	6.8	360	19	W34566
44	110.5	6.8	532	12	R13229
45	110	6.7	531	16	W01503

ALIGNMENTS

RESULT	1
Y54122	Y54122 standard; Protein; 490 AA.
XX	XX
AC	Y54122;
XX	XX
DT	27-MAR-2000 (first entry)
XX	XX
DE	Amino acid sequence of a Bacillus mannanase enzyme.
XX	XX
KW	Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW	endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; glucomannan;
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW	galactoglucomanan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW	printing paste; plant material degradation; recycled waste paper;
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;
KW	manan-containing food; coffee extract; cleaning composition;
KW	machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW	contact lens; body-care composition; fabric softener; oil well drilling;
KW	subterranean formation fracture.
XX	XX
OS	Bacillus sp.
XX	XX
FH	Key
FT	Peptide
FT	/note- "signal peptide"
FT	Domain
FT	/note- "catalytic domain"
FT	Misc-difference
FT	/note- "encoded by CTT"
FT	Domain
FT	/note- "linker"
FT	Domain
FT	/note- "domain of unknown function"
XX	XX

P300-CeIB fusion c
P300-CeIB fusion c
P300-CeIB fusion c
P300-CeIB fusion c
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
Sequence of alkali
Cellulase. Bacill
Bacillus cellulase
NK-1 cellulase. B
Actinomycete sp. 3
Teredinbacter end
Sequence encoded b
Alkaline cellulase
Fragment of alkali
Cellulase. Bacill
Full length Bacill
A. cellulolyticus
A. cellulolyticus
Acidothermus cellu
Corrected Bacillus
Globodera rostoch
Globodera glycine
Heterodera glycine
Bankia gouldi glyc
Heterodera glycine
Thermotoga OCI/4V
Thermotoga OCI/4V
Endoglucanase enco
60 kd endoglucanas

```

PN      WC0964619-A2.
XX      16-DEC-1999.
PD      10-JUN-1999; 99NO-DK00314.
XX      10-JUN-1998; 98US-0111256.
XX      20-OCT-1998; 98DK-0001340.
PR      20-OCT-1998; 98DK-0001341.
PR      28-OCT-1998; 98US-0105970.
PR      28-OCT-1998; 98US-0106054.
PR      23-DEC-1998; 98DK-0001725.
PR      05-MAR-1999; 99DK-0000306.
PR      05-MAR-1999; 99DK-0000307.
PR      05-MAR-1999; 99DK-0000308.
PR      05-MAR-1999; 99DK-0000309.
PR      09-MAR-1999; 99US-0123543.
PR      10-MAR-1999; 99US-0123623.
PR      10-MAR-1999; 99US-0123641.
PR      11-MAR-1999; 99US-0123642.
XX      (NOVO ) NOVO-NORDISK AS.
XX      Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PI      WPI; 2000-105891/09.
DR      N-PSDB; 245335.
XX      New mannanses for treatment of textiles, plant material and coffee
PT      extract, and in cleaning compositions -
PS      Claim 1; Page 208-210; 242pp; English.
XX      The present sequence represents a mannanase (also known as mannan
CC      endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
CC      of Bacillus sp. 1633. The mannanase hydrolyses galactomannans.
CC      Specifically, mannanses hydrolyse 1,4-beta-D-mannosidic linkages in
CC      mannans, galactomannans, glucomannans, and galactoglucomannans. The
CC      mannanase protein, or preparations containing it, are used to improve
CC      properties of cellulosic or synthetic fibres, yarn or (non)woven
CC      fabrics (removal of mannan-based sizes or printing pastes). They are
CC      also used to degrade or modify plant materials (particularly recycled
CC      waste paper, paper making pulp, or material containing guar or locust
CC      bean gums (thickeners), or to reduce viscosity of mannan-containing
CC      foods or feeds). The mannanses are also used to process coffee
CC      extracts (to inhibit gel formation); in cleaning compositions (for
CC      machine washing of fabrics, as hard-surface cleaners, for hand or
CC      machine dishwashing, also in oral, dental, contact lens or body-care
CC      compositions) where they remove mannan-containing soils and prevent
CC      binding of some soils to cellulosics; and in fabric softeners. They
CC      can also be used in oil well drilling to fracture subterranean
CC      formations.
CC      SQ      Sequence 490 AA:

```

```

Query Match      100.0%; Score 1630; DB 21; Length 490;
Best Local Similarity 100.0%; Pred. No 17,8e-141;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 NANSQFVSGTLLYDANGNPFVARGINHGAWYKDDATTAIEGIANTGANTVRIYVLSGG 60
DB      31 nansqfvgstlltydangnpfvmrglnhgwkykdgatacgaatgntvriyvsdg 90
QY      61 QWTKDIIHYVRNIIISLAEDNHLVAVPEVHADATGYDSIASNRAVDVWIMERSALLICKEDT 120
DB      91 qwtkdiihyvrniiislaednhlvavpevhadatydsiasnrvdvwimertallikedt 150
QY      121 VIININENWEGSWEGDAMADGYKOATPRLNAGINNTIMWDAGKQOPPOSISIDYREYF 180
DB      151 viininenwsgwedagwadgykqatprlnaglnntlmwdaagwqfpgslidyreyf 210
QY      181 MADPQRTMFSSIMYEVAGNASQVTRNTIDRVINQDLALVIGFGRHRTNGDVDEATIMS 240

```

```

DB      211 naqgrntmfslhmeyagngsqvtrntidrvlnqdlalvlgfgrhrtngdvdeatims 270
QY      241 YSEQRGVGLAWSKNGGPEWEYLDISNDAGNNLTANGNTIVNGPYGLRETSRLSTVFT 300
DB      271 yseqrgvgwlawswknggpeweyldisndwagnltawngtlvngpyglretrslstvt 330

RESULT 2
ID      Y54123 standard; Protein; 476 AA.
XX      Y54123;
AC      Y54123;
DT      27-MAR-2000 (first entry)
XX      A mannanase-linker-cellulose binding domain fusion protein.
DE      Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX      endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; glucomannan;
KW      1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW      galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW      printing paste; plant material degradation; recycled waste paper;
KW      paper making pulp; guar; locust bean gum; thickener; viscosity;
KW      mannan-containing food; coffee extract; cleaning composition;
KW      machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW      contact lens; body-care composition; fabric softener; oil well drilling;
KW      subterranean formation fracture; cellulose binding domain.
XX      Synthetic.
OS      Bacillus sp.
OS      Clostridium thermocellum.
XX      WO9964619-A2.
PD      16-DEC-1999.
XX      10-JUN-1999; 99NO-DK00314.
PE      10-JUN-1998; 98US-0111256.
PR      20-OCT-1998; 98DK-0001340.
PR      20-OCT-1998; 98DK-0001341.
PR      28-OCT-1998; 98US-0105970.
PR      28-OCT-1998; 98US-0106054.
PR      23-DEC-1998; 98DK-0001725.
PR      05-MAR-1999; 99DK-0000306.
PR      05-MAR-1999; 99DK-0000307.
PR      05-MAR-1999; 99DK-0000308.
PR      05-MAR-1999; 99DK-0000309.
PR      09-MAR-1999; 99US-0123543.
PR      10-MAR-1999; 99US-0123623.
PR      10-MAR-1999; 99US-0123641.
PR      11-MAR-1999; 99US-0123642.
XX      (NOVO ) NOVO-NORDISK AS.
XX      Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PI      WPI; 2000-105891/09.
DR      N-PSDB; 245336.
XX      New mannanses for treatment of textiles, plant material and coffee
PT      extract, and in cleaning compositions -
PS      Example 4; Page 211-212; 242pp; English.
XX      The present sequence represents a mannanase-linker-cellulose binding
CC      domain fusion protein. Mannanase (also known as mannan
CC      endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
CC      hydrolyses galactomannans. Specifically, mannanses hydrolyse
CC      1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
CC      and galactoglucomannans. The mannanase protein, or preparations
CC      containing it, are used to improve properties of cellulosic or

```

CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC celluloses; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.

SO Sequence 476 AA;

Query Match 99.6%; Score 1624; DB 21; Length 476;
Best Local Similarity 100.0%; Pred. No. 2.6e-140;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ANSGFTVSGTITLYDANGNPFVNRGINHGHWYKQDATTIEGIANTGANTVRIYISDGGQ 61
DB 1 ansqfvgstltlydangnprfvmrghnhykqdtatlaegiantgntvriyisdggq 60
OY 62 WTKDIIHYRNLIISAEQNLVAPEVHDATGYSIASINRAVDYIEMRSALIGKEDTV 121
DB 61 wkddihvtrnlislaednhlvaapevhdacgydsiasinravdyiemrsligkdv 120
OY 122 IININEMFGSWEGDAMADGYKQALPRLNAGLINTLWADAGWQGFOSIHIDYGREYFN 181
DB 121 iinlnemfsgswgdawadgykqalprlnaglnhtlmdaagwqgfpsihdygrevfn 180
OY 132 ADPQRTMFSSIMYRYAGNSQVRTNIDRYINODIALVIGFGRHRTNGVDVDEATINSY 241
DB 131 adpqrntmfsslmryagnsqvrtnidryinodlavlvgfgrhrtngdvddeatimsy 240
OY 242 SEQRGVGLAWSKGNPWEYLDLSNDWAGNNLTAKGNTYNGPGLRSTSLTVFT 300
DB 241 seqrgvglawskgnpweyldlsndwagnnltakgntlyngpplrslstlvtft 299

RESULT 3
Y44496
ID Y44496 standard; Protein: 468 AA.

AC Y44496;

DT 27-MAR-2000 (first entry)

DE Bacillus agaradherens Clone MB594, Mannanase enzyme.

KW Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KW detergent composition; mid-branched anionic surfactant; washing;
KW cosmetic stain; food stain.

OS Bacillus agaradherens.
OS Synthetic.

XX Key Location/Qualifiers
XX Peptide 1..31
XX FT /Label- Signal peptide
XX FT 32..468
XX FT Protein /Label- Mature_Mannanase

PN W09964552-A1.

PD 16-DEC-1999.

PF 10-JUN-1998; 98WO-US12026.

PR 10-JUN-1998; 98WO-US12026.

XX (PROC) PROCTER & GAMBLE CO.

XX Bettiol JP, Thoen CAJL;
PI

XX WPI: 2000-116536/10.

DR N-PSDB; 229846.

XX Detergent composition for removing greasy stains such as cosmetics,
PI food stains and body soils

XX Disclosure: Page 103-104; 113pp; English.

CC The present sequence is the B. agaradherens Clone MB594, alkaline
CC mannanase enzyme. Clone MB594 is derived from B. agaradherens strain
CC NCTMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
CC from 7.5-10.5. It can be used in a detergent composition along with a
CC mid-branched anionic surfactant. The detergent composition may be used
CC for washing purposes, soaking/pre-treatment of stained fabric, hard
CC surface cleaning and for removal of cosmetic and/or food stains. This
CC composition provides excellent cleaning effect at low temperature.

SO Sequence 468 AA;

Query Match 78.7%; Score 1283; DB 21; Length 468;
Best Local Similarity 77.3%; Pred. No. 3.5e-109;
Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

OY 2 ANSGFTVSGTITLYDANGNPFVNRGINHGHWYKQDATTIEGIANTGANTVRIYISDGGQ 61
DB 32 asqfvgstltlydangnprfvmrghnhykqdtatlaegiantgntvriyisdggq 91
OY 62 WTKDIIHYRNLIISAEQNLVAPEVHDATGYSIASINRAVDYIEMRSALIGKEDTV 121
DB 92 wkddihvtrnlislaednhlvaapevhdacgydsiasinravdyiemrsligkdv 151
OY 122 IININEMFGSWEGDAMADGYKQALPRLNAGLINTLWADAGWQGFOSIHIDYGREYFN 181
DB 152 iinlnemfsgswgdawadgykqalprlnaglnhtlmdaagwqgfpsihdygrevfn 211
OY 182 ADPQRTMFSSIMYRYAGNSQVRTNIDRYINODIALVIGFGRHRTNGVDVDEATINSY 241
DB 212 adpqrntmfsslmryagnsqvrtnidryinodlavlvgfgrhrtngdvddeatimsy 271
OY 242 SEQRGVGLAWSKGNPWEYLDLSNDWAGNNLTAKGNTYNGPGLRSTSLTVFT 300
DB 272 seqrgvglawskgnpweyldlsndwagnnltakgntlyngpplrslstlvtft 330

RESULT 4
Y54125
ID Y54125 standard; Protein: 468 AA.

AC Y54125;

DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a Bacillus mannanase enzyme.

XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture.

OS Synthetic.

OS Bacillus agaradherens.

PN W09964619-A2.

BD 16-DEC-1999.
 XX PF 10-JUN-1999; 99WO-DK00314.
 XX PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 09-MAR-1999; 99DK-0000309.
 PR 10-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 11-MAR-1999; 99US-0123641.
 XX PR 11-MAR-1999; 99US-0123642.
 XX PA (NOVO) NOVO-NORDISK AS.
 XX PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX DR WPI; 2000-105891/09.
 XX DR N-PSDB; 245338.
 XX PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions -
 XX Example 5; Page 215-216; 242pp; English.
 XX PS
 XX The present sequence represents a mannase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannase, or endo-1,4-mannase).
 CC The present mannase is a synthetic variant of the mannase of
 CC T34124, in which the C-terminus of the protein was changed due to
 CC design of a lower PCR primer used for amplification. The mannase
 CC hydrolyses galactomannans. Specifically, mannases hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 XX Sequence 468 AA;
 SQ

Query Match 78.7%; Score 1283; DB 21; Length 468;
 Best Local Similarity 77.3%; Pred. No. 3.5e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

QY 2 ANSGFYVSGTLLYDANGNPFVNRGINHGHWYKDOATTAIEGIANTGANTVRIYLSGQ 61
 DB 32 astgfyvgnltlydangpffvmrghnhykdcstalpataegantlttlylsdgq 91
 QY 62 WTKDDIHTVRNLISLAEDNHLVAPEVHDATGDSIASLNRAVDYIEMRSALIGKEDTV 121
 DB 92 wkdidditrtrevelaegkmvavvevdhdtgrdsrdlnravdyiemkdallgkedtv 151
 QY 122 IININEMGWSGDMAGYKQAIPIRLNAGINHTLMDAGMOGPOSIIDYGEVFN 181
 DB 152 IINIAEYGSWGSAGWYIDVLPKLTDAQICHLIMWDAAGWGYPSLINDYGGVDV 211
 QY 182 ADPORTMFSIMHYEYAGNASOVRTNIDRVINODLATVIGFGRHRTGVDVDEATIMSY 241

DB 212 adpikntfslmtyeyagddantvrsndrvldqalvigeifgnrthdgvdeditlsy 271
 QY 242 SEORGVGWLAWSKNGNGPEWEXIDLSNDMAGNNLTAMGNTIVNGPYGRSRLSTVFT 300
 DB 272 seetgtglawskngnstewdyldlsewagqhltdwgnrlvhgagdgqetskptvft 330

RESULT 5
 Y44495
 ID Y44495 standard; Protein; 493 AA.
 AC Y44495;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Bacillus agaradherens NCIMB 40482, Mannanase enzyme.
 XX
 KW Mannanase: endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
 KW detergent composition; mid-branched anionic surfactant; washing;
 KM cosmetic stain; food stain.
 XX
 OS Bacillus agaradherens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..32
 FT /label- Signal_peptide 33..343
 FT Protein /label- Mature_Mannanase
 FT
 PN W09964552-A1.
 XX
 PD 16-DEC-1999.
 XX
 XX 10-JUN-1998; 98WO-US12026.
 PF 10-JUN-1998; 98WO-US12026.
 XX
 PR 10-JUN-1998; 98WO-US12026.
 XX
 PA (PROC) PROCTER & GAMBLE CO.
 XX
 PI Bettiol JP, Thoen CAUK;
 XX
 DR WPI; 2000-116536/10.
 XX N-PSDB; 229845.
 XX
 PT Detergent composition for removing greasy stains such as cosmetics,
 PT food stains and body soils -
 XX
 PS Disclosure; Page 102; 113pp; English.
 XX
 CC The present sequence is the B. agaradherens NCIMB 40482, alkaline
 CC mannase enzyme. Mannanase shows maximum activity at pH ranging
 CC from 7.5-10.5. It can be used in a detergent composition along with a
 CC mid-branched anionic surfactant. The detergent composition may be used
 CC for washing purposes, soaking/pre-treatment of stained fabric, hard
 CC surface cleaning and for removal of cosmetic and/or food stains. This
 CC composition provides excellent cleaning effect at low temperature.
 XX Sequence 493 AA;
 SQ

Query Match 78.7%; Score 1283; DB 21; Length 493;
 Best Local Similarity 77.3%; Pred. No. 3.7e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

QY 2 ANSGFYVSGTLLYDANGNPFVNRGINHGHWYKDOATTAIEGIANTGANTVRIYLSGQ 61
 DB 32 astgfyvgnltlydangpffvmrghnhykdcstalpataegantlttlylsdgq 91
 QY 62 WTKDDIHTVRNLISLAEDNHLVAPEVHDATGDSIASLNRAVDYIEMRSALIGKEDTV 121
 DB 92 wkdidditrtrevelaegkmvavvevdhdtgrdsrdlnravdyiemkdallgkedtv 151


```

FT Peptide 1..32
FT /note= "signal peptide"
FT 33..331
FT Domain /note= "catalytic domain"
FT
FT MO9964619-A2.
FT
FT 16-DEC-1999.
FT
FT 10-JUN-1999; 99WO-DK00314.
FT
FT 10-JUN-1998; 98US-0111256.
FT 20-OCT-1998; 98DK-0001340.
FT 20-OCT-1998; 98DK-0001341.
FT 28-OCT-1998; 98US-0105970.
FT 28-OCT-1998; 98US-0106054.
FT 23-DEC-1998; 98DK-0001725.
FT 05-MAR-1999; 99DK-0000306.
FT 05-MAR-1999; 99DK-0000307.
FT 05-MAR-1999; 99DK-0000308.
FT 05-MAR-1999; 99DK-0000309.
FT 09-MAR-1999; 99US-0123543.
FT 10-MAR-1999; 99US-0123623.
FT 10-MAR-1999; 99US-0123641.
FT 11-MAR-1999; 99US-0123642.
FT
FT (NOVO ) NOVO-NORDISK AS.
FT
FT Kaupinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
FT WPI: 2000-105891/09.
FT N-PSDB; Z45340.
FT
FT New mannanses for treatment of textiles, plant material and coffee
FT extract, and in cleaning compositions
FT
FT Claim 35; Page 220-221; 242pp; English.
FT
FT
FT The present sequence represents a Bacillus mannanase (also known as
FT mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
FT The mannase hydrolyses galactomannans. Specifically, mannases
FT hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
FT glucomannans, and galactoglucomannans. The mannase protein, or
FT preparations containing it, are used to improve properties of cellulosic
FT or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
FT sizes or printing pastes). They are also used to degrade or modify
FT plant materials (particularly recycled waste paper, paper making pulps,
FT or material containing guar or locust bean gums (thickeners), or to
FT reduce viscosity of mannan-containing foods or feeds). The mannases
FT are also used to process coffee extracts (to inhibit gel formation); in
FT cleaning compositions (for machine washing of fabrics, as hard-surface
FT cleaners, for hand or machine dishwashing, also in oral, dental, contact
FT lens or body-care compositions) where they remove mannan-containing
FT soils and prevent binding of some soils to cellulosics; and in fabric
FT softeners. They can also be used in oil well drilling to fracture
FT subterranean formations.
FT
FT
FT Sequence 331 AA:
FT
FT
FT Query Match 69.7%; Score 1136; DB 21; Length 331;
FT Best Local Similarity 68.7%; Pred. No. 5.7e-96;
FT Matches 204; Conservative 47; Mismatches 46; Indels 0; Gaps 0;

```

```

DB 154 ininewyagwgdgavargygnalrqlrlnaglstftmvdagaygqyqsvvdygve.lna 213
OY 183 DPQRNMFSLHMYEYAGGNASQVFTNIDRYLNDLALVIGEPGRHRTNGVDENTINSYS 242
DB 214 dprntmfsvlmhyeyaggdantvrrnidsllsqnlavlgfghwbydgvdedtlisys 273
OY 243 EQRGVGLAWSKNGKPEWEYLDLSNDMAGNLTAMKNTIYNGPYGLRSTSLSTVE 299
DB 274 qqrnvglawswngnsegveyldlsndfagurltwgdrlvngpnlrqlrskssvfi 330

RESULT 8
Y54129
ID Y54129 standard; Protein; 369 AA.
AC Y54129;
XX
DT 27-MAR-2000 (first entry)
XX
DE Amino acid sequence of a Bacillus sp. AA349 mannanase enzyme.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW mannan; galactomannan; galactomannan; 1,4-beta-D-mannosidic linkage;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture.
XX
XX Bacillus sp.
XX
XX Key location/Qualifiers
XX PH 1..25
XX FT Peptide
XX FT /note= "signal peptide"
XX FT 26..67
XX FT Domain
XX FT /note= "N-terminal domain of unknown function"
XX FT 68..369
XX FT /note= "catalytic domain"
XX
XX MO9964619-A2.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1999; 99WO-DK00314.
XX
XX 10-JUN-1998; 98US-0111256.
XX 20-OCT-1998; 98DK-0001340.
XX 20-OCT-1998; 98DK-0001341.
XX 28-OCT-1998; 98US-0105970.
XX 28-OCT-1998; 98US-0106054.
XX 23-DEC-1998; 98DK-0001725.
XX 05-MAR-1999; 99DK-0000306.
XX 05-MAR-1999; 99DK-0000307.
XX 05-MAR-1999; 99DK-0000308.
XX 05-MAR-1999; 99DK-0000309.
XX 09-MAR-1999; 99US-0123543.
XX 10-MAR-1999; 99US-0123623.
XX 10-MAR-1999; 99US-0123641.
XX 11-MAR-1999; 99US-0123642.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Kaupinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX WPI: 2000-105891/09.
XX N-PSDB; Z45342.
XX
XX New mannanses for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions
XX

```


Db 147 vllanewyavrsdwaesagaprlrsaglahclivdaagwqypasibnergadvf 206
 QY 181 NADQORNTMSEIHXYEAGVAGNAGVFRNIDVLRNODALVIGEGHRTNGDVDEATIMS 240
 Db 207 asdplkntmislhmyagadratsenldgvlnealvgeghrhgdvdedalia 266
 QY 241 YSEQRGVGLWMSKNGPEWEYIDLSDNMGNNILTAGNNTIVNGPYGLRETSRL 295
 Db 267 ytaerqvgwlawswygnaggyeyldltegpgs-pltswgerliyvgemglkvldhl 320

RESULT 10
 Y54133
 ID Y54133 standard; Protein; 188 AA.
 XX
 AC Y54133;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Amino acid sequence of a partial *Bacillus* sp. mannanase enzyme.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KM endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KM mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KM plant material degradation; recycled waste paper; paper making pulp;
 KM guar; locust bean gum; mannan-containing food; coffee extract;
 KM cleaning composition; machine washing; hard-surface cleaner;
 KM dishwashing; oral; dental; contact lens; body-care composition;
 KM fabric softener; oil well drilling; subterranean formation fracture.
 XX
 OS *Bacillus* sp.
 XX
 PN W09964619-A2.
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 98US-0123543.
 PR 10-MAR-1999; 98US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauplien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI, 2000-105891/09.
 DR N-PSDB; 245346.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 XX extract, and in cleaning compositions
 PS
 PS Claim 35; Page 229-230; 242pp; English.
 CC The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based

CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing; also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 XX
 SQ Sequence 188 AA;
 XX

Query Match 32.0%; Score 522; DB 21; Length 188;
 Best Local Similarity 59.9%; Pred. No. 3e-40;
 Matches 97; Conservative 20; Mismatches 45; Indels 0; Gaps 0;

QY 1 NANGFYVSGTLYDANGNPFVARGIHGHAWYKDOATTAIEGANTGATVRIYSDGG 60
 Db 27 saqgfiwkgfclldkngdpyvmrghvhwfkgdleealpaaegantvrlvsngq 86
 QY 61 QWTKDDIHVYANLISLAEDNHLVAVPEVHDATGYDSIASINRAVDYWIENRSALIGKEDT 120
 Db 87 qwekdaseelarvtaateyglttvlevhdatsdnpdddkavdyiemadvlkgredr 146
 QY 121 VIINANWEGSWGDAMADGKQALPRLNAGINHTLWYDA 162
 Db 147 vllanewyavrsdwlwaxayagaprlrsaglahcllida 188

RESULT 11
 Y54130
 ID Y54130 standard; Protein; 305 AA.
 XX
 AC Y54130;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Amino acid sequence of a partial *Bacillus* sp. mannanase enzyme.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KM endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KM mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KM plant material degradation; recycled waste paper; paper making pulp;
 KM guar; locust bean gum; mannan-containing food; coffee extract;
 KM cleaning composition; machine washing; hard-surface cleaner;
 KM dishwashing; oral; dental; contact lens; body-care composition;
 KM fabric softener; oil well drilling; subterranean formation fracture.
 XX
 OS *Bacillus* sp.
 XX
 PN W09964619-A2.
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 98US-0123543.
 PR 10-MAR-1999; 98US-0123623.
 PR 10-MAR-1999; 99US-0123641.

PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 XX N-PSDB: 245343.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 XX
 PS Disclosure: Page 225-226; 242pp; English.
 XX
 CC The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulotics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 XX
 SQ Sequence 305 AA;

Query Match 22.6%; Score 369; DB 21; Length 305;
 Best Local Similarity 45.3%; Pred. No. 5,4e-26;
 Matches 73; Conservative 32; Mismatches 46; Indels 10; Gaps 4;

OY 146 IRLNAGLNTMVDAGQFPOSIDHYGRFVNADPQRTMFSIHMYETAGNAGSOV 205
 DB 1 IRLNAGLNTMVDAGQFPOSIDHYGRFVNADPQRTMFSIHMYETAGNAGSOV 205
 OY 206 RKNIDRVNODIALYGEFGRHRTNCD-----VDEATIMSTSEORGVCWLMASWNGCP 259
 DB 60 GELGALQGLIYVAMIGETGYNGNMLGSGVNAQELMNAQAGKGYGYMPGWTGND 119
 OY 260 EWEYDL-SNDAGNLTMTMGWTYNGPYGLRSTSLSTVF 299
 DB 120 ANSWIDMTNDW-GLTISWGNLVVNGTNGIRATSVATVF 158

RESULT 12
 Y54131
 ID Y54131 standard; Protein: 132 AA.

AC Y54131;
 DT 27-MAR-2000 (first entry)
 XX
 DE Amino acid sequence of a partial Bacillus mannanase enzyme.

XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; galactoglucomannan; galactomannans;
 KW synthetic fibres; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.

OS Bacillus clausii.
 XX
 XX W09964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1999; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 XX N-PSDB: 245344.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 XX
 PS Disclosure: Page 227; 242pp; English.
 XX
 CC The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulotics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 XX
 SQ Sequence 132 AA;

Query Match 19.5%; Score 318; DB 21; Length 132;
 Best Local Similarity 55.6%; Pred. No. 7,8e-22;
 Matches 60; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

OY 2 ANSGFYSGTILYDANGFNVRGINHGAWYKQDATTATIEGANTGATVRIVSDGQ 61
 DB 23 AGTFYGTGTLIDGEGNPNYGVNHGWSFKGLDITAPLATGANTVRIVLSNGQ 82
 OY 62 WKKDIIHYRNKISLAEDNHLVAVEVDATGYSIASNRVQWIE 109
 DB 83 WERTIVAEVRLAVLEEGTLAVLVDAATGSDPNDLITAVEVSE 130

RESULT 13
 W12378
 ID W12378 standard; Protein: 410 AA.

```

XX AC W12378;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 1 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. M4 (ATCC 21833).
XX FH Key
XX FT 1..26 Location/Qualifiers
XX FT /label= Sig_peptide
XX FT /note= "Hybrid between P300 (aa1-12) and
XX FT CelB (aa13-26) signal peptides"
XX FT Protein
XX FT 27..410
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN WO9706181-A1.
XX PD 20-FEB-1997.
XX PE 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENKEL) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB; T63226.
XX PT System for increased expression of cellulase and xylanase in
XX PT Bacillus - contains gene under control of elements from B.
XX PT licheniformis alkaline protease gene
XX PS Disclosure: Fig 7,7A,7B; 37pp; English.
XX CC The polypeptide product (W12378) of P300-CelB fusion construct 1
XX CC (T63226) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. M4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also
XX CC W12379-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 410 AA;

Query Match 12.4%; Score 202; DB 18; Length 410;
Best Local Similarity 25.0%; Pred. No. 1.4e-10;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

OY 8 VSGTLYDNGNPFVARGI-NHGAWYKDOATPAIEGIANVTAVRYIVLSGGQMTKDD 66
DB 41 isngelvdndrgepvqikmshnglqygy-----qfnyesmknwlrdd 82
OY 67 -----IHT-----VRNLISLAEDNHLVAPVHDATGYDSIASLNR 103
DB 83 wgtlvtiraamytsaggyiedpsvkexveasaidlglyvldwhllednphnykeea 142
OY 104 VDWIMRSALIKEDTVIININWFGS---WEGD-AMADGYKQAIPLRNAGINHTL 158
DB 143 ktdidem-selygdvnyvlyelanepgsdvtvndqikpyae---evlpylrndpnll 198

```

```

OY 159 MYDAAGMGQFPQSIDYGRVFNADPQRTKMSIMHREYAGNASQVRNIDRVNODLA 218
DB 199 iv---gtgtwsgdvh-haadngltcp--nvmvafhf--yagthgqnldqvayldgaa 250
OY 219 LVIGFEGHRTNGD-----VDEATI-MSYSEORGVGWLMAS 253
DB 251 lfsewgtseatgdggyvfideagvldfndernlswans 290

RESULT 14
W12379
ID W12379 standard; Protein; 411 AA.
XX AC W12379;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 2 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. M4 (ATCC 21833).
XX FH Key
XX FT 1..26 Location/Qualifiers
XX FT /label= Sig_peptide
XX FT /note= "Hybrid between P300 (aa1-5) and
XX FT CelB (aa8-26) signal peptides"
XX FT Protein
XX FT 27..411
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN WO9706181-A1.
XX PD 20-FEB-1997.
XX PE 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENKEL) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB; T63227.
XX PT System for increased expression of cellulase and xylanase in
XX PT Bacillus - contains gene under control of elements from B.
XX PT licheniformis alkaline protease gene
XX PS Disclosure: Fig 8,8A,8B; 37pp; English.
XX CC The polypeptide product (W12379) of P300-CelB fusion construct 2
XX CC (T63227) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. M4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also W12378,
XX CC W12380-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 411 AA;

Query Match 12.4%; Score 202; DB 18; Length 411;
Best Local Similarity 25.0%; Pred. No. 1.4e-10;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:06:50 ; Search time 154.68 Seconds
(Without alignments)
32.508 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 NANSGRFYSGTLYDANGNP.....TIVNGPYGLRHSRLSTVFT 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	12.3	400	US-08-713-298B-2	Sequence 2, Appl1
2	201	12.3	400	US-08-870-180B-2	Sequence 2, Appl1
3	201	12.3	400	US-08-814-052-4	Sequence 4, Appl1
4	201	12.3	400	US-08-812-829-4	Sequence 4, Appl1
5	201	12.3	462	US-08-870-180B-13	Sequence 13, Appl1
6	191	11.7	467	US-08-727-548-2	Sequence 2, Appl1
7	171.5	10.5	430	US-08-924-440-2	Sequence 2, Appl1
8	142	8.7	358	US-08-913B-11	Sequence 11, Appl1
9	142	8.7	521	US-08-276-213-3	Sequence 3, Appl1
10	135	8.7	551	US-09-033-537A-1	Sequence 1, Appl1
11	110	6.7	531	US-07-862-588B-7	Sequence 7, Appl1
12	100	6.1	429	US-08-745-977-4	Sequence 4, Appl1
13	100	6.1	429	US-09-040-699A-4	Sequence 4, Appl1
14	99.5	6.1	317	US-09-066-075-2	Sequence 2, Appl1
15	99.5	6.1	317	US-08-518-615A-2	Sequence 2, Appl1
16	99.5	6.1	317	US-08-951-889-2	Sequence 2, Appl1
17	97.5	6.0	574	US-08-732-433-1	Sequence 1, Appl1
18	95.5	5.9	735	US-08-313-185-48	Sequence 48, Appl1
19	95.5	5.9	735	US-08-459-499-9	Sequence 9, Appl1
20	94	5.8	398	US-08-853-659A-45	Sequence 45, Appl1
21	93.5	5.7	566	US-07-862-588B-4	Sequence 4, Appl1
22	90.5	5.6	735	US-08-459-499-12	Sequence 12, Appl1
23	88	5.4	531	US-08-531-601-1	Sequence 1, Appl1
24	88	5.4	531	US-08-859-032-1	Sequence 1, Appl1
25	88	5.4	535	US-08-737-597-10	Sequence 10, Appl1
26	86	5.3	461	US-08-672-571A-3	Sequence 3, Appl1
27	86	5.3	490	US-08-672-571A-1	Sequence 1, Appl1
28	85.5	5.2	385	US-08-387-942C-26	Sequence 26, Appl1

29	85.5	5.2	997	US-08-387-942C-4	Sequence 4, Appl1
30	85.5	5.2	1248	US-08-348-353-17	Sequence 17, Appl1
31	85.5	5.2	1248	US-08-465-965-17	Sequence 17, Appl1
32	85.5	5.2	1248	US-08-465-966-17	Sequence 1, Appl1
33	85	5.2	881	US-08-333-901-1	Sequence 1, Appl1
34	85	5.2	881	US-08-456-582-1	Sequence 1, Appl1
35	85	5.2	881	US-08-898-789-1	Sequence 1, Appl1
36	85	5.2	881	US-09-039-555B-16	Sequence 16, Appl1
37	85	5.2	4944	US-08-469-486-52	Sequence 52, Appl1
38	85	5.2	4544	US-08-469-558-52	Sequence 52, Appl1
39	84.5	5.2	2628	US-08-570-311-14	Sequence 14, Appl1
40	84	5.2	3724	US-08-804-227C-10	Sequence 10, Appl1
41	84	5.2	3724	US-08-804-198-4	Sequence 4, Appl1
42	83	5.1	514	US-08-560-656-3	Sequence 3, Appl1
43	83	5.1	531	US-08-688-988-39	Sequence 39, Appl1
44	82	5.0	455	US-08-870-827-3	Sequence 3, Appl1
45	81.5	5.0	557	US-08-793-229-33	Sequence 33, Appl1

ALIGNMENTS

RESULT 1
US-08-713-298B-2
Sequence 2, Application US/08713298B
Patent No. 5922586
GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dammann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Sch lein, Martin
APPLICANT: J rgensen, Per
APPLICANT: Bjcernevad, Mads
TITLE OF INVENTION: DNA Constructs and Methods of Producing
TITLE OF INVENTION: Cellulytic Enzymes
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5922586 No. 5922586disk of No. 5922586th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,298B
FILING DATE: 13-SEPT-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valetta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3794,424-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 867-0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-298B-2

Query Match 12.3% Score 201; DB 2; Length 400;
Best Local Similarity 26.6% Pred. No. 2e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

```

QY      8 VSGTLLYDANANPEFMRI -NHGAAV-----KDAQATALEGIANGANVRI 54
Db      40 ISNGELVNEREDQOLKMSHGLQMGQFVNYSKMRD-----DWGINFRA 89
QY      55 VL -SDGQWTKDDI-HTVRNLISLADBNHLVAAPVEYHDAATGYSIASLNRAVDYIEMR 111
Db      90 AMYTSGGYIDDPVKEKFKVAEALADLDIYVILDMHILSDNPNRYKEEAKDFDEM -148
QY      112 SALTGKEDTVLIINIAEMFGSMBEDAAADYK-----CAIPRLRAGLNHLMDADMGQ 167
Db      149 SELIGDIYRVIIETAEANEKS--DYVMGNQIKPYAEVYIPIIRNNDPNNITIV--GTGT 203
QY      168 FPOSIIHDYGREVENADPORMTFSIHMYEYAGNASOVRKNIDRVLNQDIALYIGEGHR 227
Db      204 WSDVHV-HAADNQLADP-NVMYAFHF--YAGTGQMLRQDQVDYALDQGAALYSENGTS 258
QY      228 HTNGD----VDEATI-KSYSPQBGVGLAWS 253
Db      259 AATGGGCVFLDEAQWVIDFERMNLISVANNMS 289

```

RESULT 2
 US-08-870-180B-2
 Sequence 2, Application US/08870180B
 Patent No. 5945327
 GENERAL INFORMATION:
 APPLICANT: Outtrup, Helle
 APPLICANT: Damdbmann, Claus
 APPLICANT: Olsen, Arne
 APPLICANT: Blag rd-Frantzen, Henrik
 APPLICANT: Sch lein, Martin
 APPLICANT: J rgensen, Per
 APPLICANT: Bjoernvad, Mads
 TITLE OF INVENTION: DNA Constructs and Methods of Producing
 TITLE OF INVENTION: Cellulytic Enzymes
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NO. 5945327/2 No. 5945327/disk of No. 5945327/1 America, Inc
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/870,180B
 FILING DATE: 6-June-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rozek, Carol E.
 REGISTRATION NUMBER: 36,993
 REFERENCE/DOCKET NUMBER: 3794.434-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 867 0298
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 400 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-870-180B-2

Query Match	12.3%	Score 201;	DB 2;	Length 400;
Best Local Similarity	26.6%	Pred. No. 2e-11;		
Matches 72;	Conservative 44;	Mismatches 109;	Indels 46;	Gaps 14;
OY	8	VSCTLLIYANGNPFPVAKRI-NHGHWY-----KQDATTAAIEGIANGTANTVARI	54	

```

Db      40  ISNGELVNERBEOYQJOKGMSHGQWQVFNYESMKRLAD-----DMGINVFRA 89
QY      55  VL--SDGQWTKDI-HTVRNLISLADNHLVAVPEVHADATGIDSLASLRAYDIWEMR 111
Db      90  AMYTSSGGYIDDPVSKEKYEKVAEAAIDLITVIMHILISDDNPNTYKEAADFDEM- 148
QY     112  SALIGKETVTLIINAMFEWSMEDAWADGK-----CAIPLRNAGLNHTLNAVDAWGGO 167
Db     149  SELGDEYPNVYEIETANEPNGS--DVTGNGOIKPFAEEVPIIRINPNNTIIV---GTGT 203
QY     168  FPOSJHDYGRVFNADPORNMTFSIHXYEAGNASQVFNTRIDRVLNODLALTYGEGHR 227
Db     204  WSDVH-HAANDQNLADP--NMYAFHF--YAGTHGQYLROQVYALDQGAITVSENGTS 258
QY     228  HTNGD---VDEATI-MSYSRQRGVGLANS 253
Db     259  AATGGGVLEDAQWYIDFMDERLSTWANS 289

```

RESULT 3
 US-08-814-052-4
 Sequence 4, Application US/08814052
 Patent No. 6015783
 GENERAL INFORMATION:
 APPLICANT: von der Osten, Claus
 APPLICANT: Cherry, Joel R.
 APPLICANT: Bjornvad, Mads E.
 APPLICANT: Wind, Jesper
 APPLICANT: Rasmussen, Michael Dolberg
 TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
 TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
 NUMBER OF SEQUENCES: 55
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 60157830 No. 6015783disk of No. 6015783th America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/814,052
 FILING DATE: 06-MAR-1997
 CLASSIFICATION: 510
 ATTORNEY/AGENT INFORMATION:
 NAME: Lambiris, Elias J
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 4684.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TEXT:
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 400 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-814-052-4

Query Match	12.3%	Score 201	DB 3	Length 400
Best Local Similarity	26.6%	Pred. No. 2e-11		
Matches	72	Conservative	44	Mismatches 109
				Indels 46
				Gaps 14
OY	8	VS	GTT	LYDANGNP
				YMGRI-NHGHAWY-----KQ
				QATTAL
				EGIAN
				TGANTVRI
				54

DB 40 ISNGELVNERGEVOLKSGMSHGLQWYGFVNYESMKMLD-----DWGINVFR 89
QY 55 VL--SDGGQWTKDDI-HTVRNLISLAEDNHLVAVPEVHDATGYDSIASLRANDVIEMR 111
DB 90 AMTSSGGYIDDPSEVKEKKEAVEAIDDIYIIMHLSNDPPIYKEADPFDEM-148
QY 112 SALIGREDYIINIANEMWEGMGADQYK---QAIPRLNAGLNHTLWDAAGWQ 167
DB 149 SELGDPYVITYEIANEPGNS--DYTWGNQIKRYAEVPIIINNPNNTIIV---GTGT 203
QY 168 FPOSIDYGREVFNDPQRTNFSIMTEYAGNASQVNTNDVNLQDALVIGEGFR 227
DB 204 WSGDVA-HAADNQLADP--NVMTAFHF--YAGTHGONLRDQVYALDQGAALFVSEMGTS 258
QY 228 HTNGD----VDEATI-MSYSEORGWGLMS 253
DB 259 AATGCGVFLDEAQAVIDFDERNLMSMANS 289

RESULT 4

US-08-812-829-4
; Sequence 4, Application US/08812829
; Patent No. 6017751
; GENERAL INFORMATION:
; APPLICANT: von der Osten, Claus
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Vind, Jesper
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
; TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6017751st of No. 6017751th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/812,829
; FILING DATE: 06-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4690.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-812-829-4

Query Match 12.3%; Score 201; DB 3; Length 400;
Best Local Similarity 26.6%; Pred. No. 26-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;
QY 8 VSGTTLIDANGNDFVNRGI-NHGHANY-----KDAQTTAIEGIANTGANTVRI 54
DB 40 ISNGELVNERGEVOLKSGMSHGLQWYGFVNYESMKMLD-----DWGINVFR 89

QY 55 VL--SDGGQWTKDDI-HTVRNLISLAEDNHLVAVPEVHDATGYDSIASLRANDVIEMR 111
DB 90 AMTSSGGYIDDPSEVKEKKEAVEAIDDIYIIMHLSNDPPIYKEADPFDEM-148
QY 112 SALIGREDYIINIANEMWEGMGADQYK---QAIPRLNAGLNHTLWDAAGWQ 167
DB 149 SELGDPYVITYEIANEPGNS--DYTWGNQIKRYAEVPIIINNPNNTIIV---GTGT 203
QY 168 FPOSIDYGREVFNDPQRTNFSIMTEYAGNASQVNTNDVNLQDALVIGEGFR 227
DB 204 WSGDVA-HAADNQLADP--NVMTAFHF--YAGTHGONLRDQVYALDQGAALFVSEMGTS 258
QY 228 HTNGD----VDEATI-MSYSEORGWGLMS 253
DB 259 AATGCGVFLDEAQAVIDFDERNLMSMANS 289

RESULT 5

US-08-870-180B-13
; Sequence 13, Application US/08870180B
; Patent No. 5945327
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Dabmann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg, rd-Frantzen, Henrik
; APPLICANT: Sch, Jeln, Martin
; APPLICANT: J rgensen, Per
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; TITLE OF INVENTION: Cellulytic Enzymes
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5945327th of No. 5945327th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,180B
; FILING DATE: 6-June-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 3794.434-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-870-180B-13

Query Match 12.3%; Score 201; DB 2; Length 462;
Best Local Similarity 26.6%; Pred. No. 2-5e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;
QY 8 VSGTTLIDANGNDFVNRGI-NHGHANY-----KDAQTTAIEGIANTGANTVRI 54
DB 40 ISNGELVNERGEVOLKSGMSHGLQWYGFVNYESMKMLD-----DWGINVFR 89
QY 55 VL--SDGGQWTKDDI-HTVRNLISLAEDNHLVAVPEVHDATGYDSIASLRANDVIEMR 111

Db 90 AMTSSSGYIDDPVSKKVEAEVAEADLDIYIIDWHILISDNDPNYREADDFPEM- 148
QY 112 SALIGKEDYIINIANFWGWSWEGDANADYK---QALPRLNAGLNTLAWDAAGWGQ 167
Db 149 SELTGDYPNVYIEIANEPNS--DVTWNGQIKPFAEEVPIIRNDNNIITIV---GTG 203
QY 168 FPOSIDHGREVFNADQQRMTFSIHMYEYAGGASQVFNINIDRVLMODLALVIGEGHR 227
Db 204 WSDQVH-HAADNOLADP--NMYAFHF--YAGTHGQNLROQVADLQGAIFVSEWGT 258
QY 228 HINGD---VDEATI-MSYSEORGVGWLANS 253
Db 259 AATGDGGVFLEDAQVWIDFMDERLISWANN 289

RESULT 6
US-08-727-548-2
; Sequence 2, Application US/08727548
; Patent No. 5856165
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: No. 5856165el Alkaline Cellulase and Methods of
; TITLE OF INVENTION: Producing Same
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,548
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-727-548-2

Query Match 11.7%; Score 191; DB 2; Length 467;
Best Local Similarity 24.6%; Pred. No. 2.3e-10;
Matches 69; Conservative 49; Mismatches 98; Indels 64; Gaps 14;
QY 8 VSGTTLTDANGNPFVWNGT-NHGHAWTKDQATRAISGIANFTAVRIVLSDGGQWTKD 66
Db 40 ISNGELVNERGEQVQLKSGMSHGIQWYG-----QFVYSESKWLDD 81
QY 67 -----IHT-----VRNLISLAEQNHLYAVPEVDANGYDSIASLNR 103
Db 82 WGTIVFRANNTSSGGYIDDPVSKKVEAEVAEADLDIYIIDWHILISDNDPNYRE 141
QY 104 VDIWIEKRSALIKEDYIINIANFWGWSWEGD--AWADGKQALPRLNAGLNTL 158
Db 142 KDFEDEM-SELYGDYPNVYIEIANEPNSGSDVTDNOKPYAE--EVIPIYIRNDNNIY 197
QY 159 MYDAAGGQFPOSIDHGREVFNADQQRMTFSIHMYEYAGGASQVFNINIDRVLMODLA 218
Db 198 IV---GTGWSQVH-HAADNOLADP--NMYAFHF--YAGTHGQNLROQVADLQGA 249
QY 219 LVIGEGHRHTNGD---VDEATI-MSYSEORGVGWLANS 253
Db 250 IFVSEWGTSAATGDGGVFLEDAQVWIDFMDERLISWANN 289

RESULT 7
US-08-924-440-2
; Sequence 2, Application US/08924440
; Patent No. 5871550
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.

APPLICANT: Frits et al.
; TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,440
; FILING DATE: August 27, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Christopher L. Stone
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC388
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 846-7555
; TELEFAX: (650) 845-6504
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-924-440-2

Query Match 10.5%; Score 171.5; DB 2; Length 430;
Best Local Similarity 24.4%; Pred. No. 1.5e-08;
Matches 77; Conservative 49; Mismatches 122; Indels 67; Gaps 18;
QY 8 VSGTTLTDANGNPFVWNGT-NHGHAWTKDQATRAISGIANFTAVRIVLSDGGQ 61
Db 137 VCGICDDEHGNVQLKSGMSHGIQWYFCHLDSLDALAYWKAIDIRLSYIQEDGE 196
QY 62 WT-----KDIHIVRNLISLAEQNHLYAVPEVDANGYDSIASLNRADVWIMRSALIG 117
Db 197 INPGEFTDPMH--OLIDATARGLYVIYDWHILTPGDPHYMLDRAKTFFAIAORHASK 253
QY 118 EDTVITINIANFWGWSWEGDAMAD--GY-KOALPRLNAGLNTLAWDAAGQFPOSIDH 174
Db 254 TN-VLKEIANE-----PNQVSNASISYAEVYIPVIRQDPPSVIIVGRGMSLQVS-EG 307
QY 175 YGREVFNADP--QORTMFSIHMYEYAGGASQVFNINIDRVLMODLAL--VLIGEGHRHT 229
Db 308 SGPAAEIAAPVNASIMYAFHF-----ASHRDNVYNALREASELPPVFYTERGETEY 361
QY 230 NND-----VDEATIMSGSEORGVGWLANSWK-----GNQPEWETLDS 267
Db 362 TGDGANDPQMDR--YIDLMERKIGTKWNSDPSFGAVFQPGTCASGP----- 411
QY 268 NDWAGNNLTAMGNTI 282
Db 412 --WSSGSLKASGQW 424

RESULT 8
US-08-604-913B-11
; Sequence 11, Application US/08604913B
; Patent No. 5712142
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.

APPLICANT: Himmel, Michael E.
APPLICANT: Baker, John O.
TITLE OF INVENTION: METHOD FOR INCREASING
TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: U.S.A.
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II (DOS) text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,913B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,213
FILING DATE: 15-070-1994
ATTORNEY/AGENT INFORMATION:
NAME: Edna M. O'Connor
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: 95-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/384-7573
TELEFAX: 303/384-7499
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FEATURE:
NAME/KEY: E1-CAT
US-08-604-913B-11

Query Match 8.7%; Score 142; DB 1; Length 358;
Best Local Similarity 23.1%; Pred. No. 7.8e-06;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

2 ANSGF-YVSGTLYDANGNPFVNRGIN-----HGHWKDKQATTAIEBIANTGA 49
1 AGGGYWTSGREILDANNVPRAGINWFGFETCNVYHG-LMSRDY-RSMLDQIKSLGY 58
50 NTVAVIVSDSGOWTKDID---HTVRNLISLAEDNHLVAVPEVHATGYDSIASINRAVDY 106
59 NTRILRPLYS-----DILRPGITPNSINFYQNN-----QDLQGLTSLQVYDKIVAY 103
107 -----WIENRSALI-----GKEDVTIINIANEW 129
104 AGQIGLRILIDRRHPCSGSALMTYSSEATWISDLQALQRYKGNPTVVGFDLHNE- 162
130 FGSWEGDAMADG-----YKQAIPLRNA--GLNHTLAV-----DAAGW----- 165
163 --PHDPACWCGGDSIDWRLAERAGNAVLVSVPNLLIFEGVQSYNGDSYMWGNTLOGA 220
166 GQFP-----QSITHDGREVFNDPQRTWFSIHMYEYAGNNSQV-RTNIDRYLN 214
221 GQYPVVLANVNRILVYSADHATISY---PQ--TWFSPTFP---NNPGLMKNKMGILFN 272
215 QDLALV-IGERGH--RRTNGVDDEATIMSY---SEQRGVG---WLANSWKGNG----- 258
273 QNLAPVWIGEGTILQSTIDTQWTKLTVQYLRPRAYGADSFQWTFVSWNDSDGTGGIL 332
259 -PEWEYLDLSND 269

Db 333 KDDWQIVDVKD 344

RESULT 9
US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 8.7%; Score 142; DB 1; Length 521;
Best Local Similarity 23.1%; Pred. No. 1.3e-05;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

2 ANSGF-YVSGTLYDANGNPFVNRGIN-----HGHWKDKQATTAIEBIANTGA 49
1 AGGGYWTSGREILDANNVPRAGINWFGFETCNVYHG-LMSRDY-RSMLDQIKSLGY 58
50 NTVAVIVSDSGOWTKDID---HTVRNLISLAEDNHLVAVPEVHATGYDSIASINRAVDY 106
59 NTRILRPLYS-----DILRPGITPNSINFYQNN-----QDLQGLTSLQVYDKIVAY 103
107 -----WIENRSALI-----GKEDVTIINIANEW 129
104 AGQIGLRILIDRRHPCSGSALMTYSSEATWISDLQALQRYKGNPTVVGFDLHNE- 162
130 FGSWEGDAMADG-----YKQAIPLRNA--GLNHTLAV-----DAAGW----- 165
163 --PHDPACWCGGDSIDWRLAERAGNAVLVSVPNLLIFEGVQSYNGDSYMWGNTLOGA 220
166 GQFP-----QSITHDGREVFNDPQRTWFSIHMYEYAGNNSQV-RTNIDRYLN 214
221 GQYPVVLANVNRILVYSADHATISY---PQ--TWFSPTFP---NNPGLMKNKMGILFN 272

215 ODLALV-IGERGH--RHTNGDVDEATIMSY---SEORGVG---WLAWSKNGKNG----- 258
Db 273 ONIAPVWIGEGGTIGSTTDTWTAKTLYQYLRPAQYAGADSEFWTFWMSWNPDSDDTGGIL 332
QY 259 -PEWEYLDLSND 269
Db 333 KDDWQTVDTVRD 344

RESULT 10
US-09-033-537A-1
Sequence 1, Application US/09033537A
Patent No. 5958083
GENERAL INFORMATION:
APPLICANT: Onishi, Masahiro
APPLICANT: Fich, Merete
APPLICANT: Toft, Annette Hanne
APPLICANT: Sh lein, Martin
TITLE OF INVENTION: Prevention Of Back-Staining
TITLE OF INVENTION: In Stone Washing
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59580830 No. 59580830th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033.537A
FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-033-537A-1

Query Match 8.3%; Score 135; DB 2; Length 551;
Best Local Similarity 20.9%; Pred. No. 6.8e-05;
Matches 81; Conservative 49; Mismatches 133; Indels 124; Gaps 19;

QY 8 VSGTLLYDANGNPFVARGI-NHGHANYKDOATTAIGTANTGANTVRIYLSGGQNTKDD 66
Db 11 VQGNQAVGSGQAVQAVLWGMSSHGLQWYGN-----FVKKSLQWMDRN 52
QY 67 --IHTVRLISLAEDNHLVAVPEVHDATGDTASLNR-----VDYWI----- 108
Db 53 WGINVRAATYTAEDY-TIDPSVKKKVKAQVQASIDLGILYIIDWHILSDGNPNTYKQ 111
QY 109 -----EMRSALIGKEDTVIINIANEWFGSWEGDAMAD--GY-KQALPRLNAGLNHTLWV 160

Db 112 SKAFQEMATLYENTPVIETIANEPNGN---VSMADVSYAEVITATRAIPDDGVIV 168
QY 161 DAAGWQFQPSIDYGREYFNADPQRTMFSIMYREAGGNSQVFTNIDRYLNDLALV 220
Db 169 GSPITWQ-----DIHLADNPVSHSNVYALHF--YSGTHGQFLHDIRLYANNGKAAIF 220
QY 221 IGEGERHTNGD-----VDEATIMSYSEORGVGWLWS-----WKNGPEWEYL 264
Db 221 VTEWGTSDASNGGPPYFQSKWIDFLNARKISWVMSLADKVTSAALMFGASPTGGWT 280
QY 265 DL-----SNDW-----AGN-----NLTAW-GN----- 280
Db 281 DAQISESGKMYRQIRKATGGGSGNPTAPAPNTLSTAGNAGVSLTWNAVSGATSYTVK 340
QY 281 -TIVNGPY-----GLNETSRSTVFT 300
Db 341 RATSGGPTNVATGVATSYTNGLT 367

RESULT 11
US-07-862-588B-7
Sequence 7, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linnae
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 59167960th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862.588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212.867.0298
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-862-588B-7

Query Match 6.7%; Score 110; DB 2; Length 531;
Best Local Similarity 19.6%; Pred. No. 0.016;
Matches 54; Conservative 45; Mismatches 118; Indels 58; Gaps 10;

QY 8 VSGTLLYDANGNPFVARGI-NHGHANYKDOATTAIGTANTGANTVRIYLSGGQNTKDD 66

Db 44 VQGNOLVQSGOAVOLVMSHGLWYCN-----FYKSSLOWMEDN 85
QY 67 --HTVNLILAEDELNLVAVPEVDATGYSIASLNA-----VDYWT----- 108
Db 86 WGINFRAAMTSESDGY-ITDPSYKKNKVEANQASIDLALYIIDMHLISDGNPTYKAQ 144
QY 109 -----EMASALIGKEDYIININANWFGSWGDANADGKALPRLRNAGLNHTLMDAA 163
Db 145 SKAFQENATLYGNTPNVITYELATSPTECVLGRCSS--EEVITAIRSIDPQGVVIVASP 202
QY 164 GWCQFPOSIDHYGREVFENADPQRTMFSIHMYEYAGNACVYRTNIDRYLNODLALVIGE 223
Db 203 TWSQ-----DHLAADNPVSHSNVYALHF--YSGTHGQFLRDXITTAAMKGAALFYTE 254
QY 224 FGHRTNGD---VDENATMSYSEQRGVGLAWS 253
Db 255 WGTSDASGNGGYPPLQSKEMIDFLNARKISWYMS 289

RESULT 12
US-08-745-977-4
; Sequence 4, Application US/08745977
; Patent No. 5770406
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglycanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5770406 No. 5770406disk of No. 5770406th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745,977
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4175,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-745-977-4

Query Match 6.1%; Score 100; DB 1; Length 429;
Best Local Similarity 21.9%; Pred. No. 0.11; Indels 88; Gaps 18;
Matches 68; Conservative 35; Mismatches 119; Indels 88; Gaps 18;
QY 27 NHGAWTKDQATTAIEGANTGANTVRIVLSDGGQWTKDIDHTVYN----- 72
Db 106 NHYDWMIN---PATVQSHVDGLMTIRIPI---GYWSNALVDNRSEFPADGNQLPFLD 159

QY 73 -LISLAEDNELVAVPEVDATG---YDSIASLN-----RAVDY-----WIEMSA 113
Db 160 AVYORRADLGIYIYIDLDGAGGQODAFTCGNPNPAGFYNTYDYGRAEKLMSWTNIIH 219
QY 114 LIGKEDTV-IININANWFGSWGDAM--ADG-----YKQAIPLRL-----NAGLN 155
Db 220 TNPASTYGVKLEVLNEPYSRHDGGRYRAPGODDSMVUTYIPGALKAVRDAEALNVP SN 279
QY 156 HTLWDAAGKQFPOSIDHYGREVFENADPQRTMFSIHMYEYAG-----GNASQVR--- 206
Db 280 KKLHY-----QEMSSKNDSDGPRSNAAVKNDPVYGFDDHNYIGFALQNTGDQSLMISA 333
QY 207 -TNIDRYLNODLALYIGFGRHRTNG-DVDEATIMS-----YSQRGVGLAWSWK 255
Db 334 CDSRVVVGQDFA-ITGEMS--MTSGADWHGDNFTKFTTAQOQLYESPGMDGMYWTWK 390
QY 256 G--NGEMEX 263
Db 391 TELNDPNTY 400

RESULT 13
US-09-040-699A-4
; Sequence 4, Application US/09040699A
; Patent No. 6022723
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglycanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 6022723 No. 6022723disk of No. 6022723th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,699A
; FILING DATE: 18-March-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4175,214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-040-699A-4

Query Match 6.1%; Score 100; DB 3; Length 429;
Best Local Similarity 21.9%; Pred. No. 0.11; Indels 88; Gaps 16;
Matches 68; Conservative 35; Mismatches 119; Indels 88; Gaps 16;
QY 27 NHGAWTKDQATTAIEGANTGANTVRIVLSDGGQWTKDIDHTVYN----- 72

Db 106 NHYDWIN---PATVQSVHDVGLNTIRPI---GYMSYNAIVDTASEPADGNLTQLYLD 159
QY 73 -LISLAEDNHLVAVPEVDATG---YDSIASLN-----RAYDY-----WIEBBSA 113
Db 160 AVVQKAAADLGIVYIIDLGAAGGGOODAFITGONPAGFYNTYDYGRAEKYLSWMTNIIH 219
QY 114 LIGKEDVY-IINIANEMGSEWDAM--ADG-----YKQAIPLRL-----NAGLN 155
Db 220 TNPATSTYGMLEVLNEPYSRHDGGRYAPGQDSBMQTYTPGALKAVRDALALNPSPN 279
QY 156 HTLWDAAGWGFQPSIDHYREYFNADPQNTMFSIHMTYAG-----GNASQVR--- 206
Db 280 KKLHV-----QFMSSKMDSGDPRSNAAVKNDPMVGFDDHNYIGFALONTGDOYSLMBSA 333
QY 207 -TNDIVYANODLALVIGFGRHNTG--DYDEATIMS-----YSRQGVGWLAMSK 255
Db 334 CTDSRVVSGQDFA-ITGEMS--MTSGADWHDGNEFTTKFTTAQOQLYESPGMDGWTYTWK 390
QY 256 G--NGPEWEY 263
Db 391 TELNDPRWTY 400

RESULT 14
US-09-066-075-2
; Sequence 2, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-066-075-2

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

QY 19 NPFVAMGINSHEGA-----W-----YKQATRAIEGANTGANTVRYVLSDGQWTK--- 64
Db 9 NKLIGRINIGNALLEAPNEDGQVVKDE---FFDIKRGSSHVPIPI---RMSSTAY 61
QY 65 -----DDI--HTVRNLISLAEDNR-----LVAVPEVDATGYSIASINRA 103
Db 62 APPPYKIMDFEKRVDVINGALKRGLAVAINHHEELMNDPEERK-----ERF 111
QY 104 VDYIEMRSALIGKEDVTIINIANEMGSEWDAMADYKQAIPLRLNAGLNTLWADAA 163
Db 112 LALMKQADRYKYPETLFEELNEPHGNTLPEKRWELLEALKVRSIDKRTIIIGTA 171
QY 164 GWGQFPQ-----SIHDYREYFNADPQNTMFSIHMT---EYAGNMQVNTNIDRYLND 216
Db 172 EWGGSIALEKISVPKW-----EKNSIVTIIHYNPFETTHOGAEWVGS--EKLIGRK 221
QY 217 LA-----LVIGFGRHNTGVDV-----ATMSYSEQRG 246
Db 222 WGSDDQKHLIEENFEIEMSKKKRPIYIGERG-AVRKADLESRLKWTGFVYREMEKR- 279
QY 247 VGWLAMSKNGPEWEY 263
Db 280 ---RMSW---AYWEF 288

RESULT 15
US-08-518-615A-2
; Sequence 2, Application US/08518615A
; Patent No. 5962258
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615A
; FILING DATE: August 23, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-518-615A-2

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

```

0Y 19 NPEVARGINBHA-----W-----YKDOATTAEGTANGLMYRVLSDGOWTK--- 64
Db 9 NKILRGGINISNALEAPNEGDMGVYIDE--FFDILKEGFSFVRIPi---RWSTHAY 61
0Y 65 -----DDI--HTVNLISLAEDNN-----LVAPEVPHATGYDSIASLNRA 103
Db 62 AEPFYKINDREFKRYDEVYNALRGGLAVAINIHIEYELMNDPEHR-----ERG 111
0Y 104 VDYWIEKMSALIGEDVYIINIANEMFGSWEGDAMADGYQOAIPLRLNAGLNTLWYDA 1633
Db 112 LALWQIADRYKOYPERLFEFLINEPHGNLPEKWMNLEBALKVINSIDKKTIIIGTA 171
0Y 164 GNGGPPQ---SINDYGHETVFNADPQRNTWFSIMT---EYAGNMSQVFTNIDRYLND 216
Db 172 EMGGISALEKLSVDRW-----EKNSIYTIYHNPFETHOGALEWYBGS--EKWLGRK 221
0Y 217 LA-----LVIGEPHRTNGDYDE-----ATMSYSQRG 246
Db 222 WGSPPDOQHLLIEFNFIEMSKKKRNPRIYIGEG-AVRKADLESRLKMTISFVYREMEKR- 279
0Y 247 VGNLAMSVMKNGPEWEX 263
Db 280 -----RWSTH---ATWEX 288

```

Search completed: December 19, 2000, 16:06:53
Job time: 395 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:10:59 ; Search time 245.85 Seconds
(without alignments)
77.438 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 NANSGFYVSGTLVDANGNP.....TIVNGPYGLRHSRLSTVPT 300

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.65:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1169	71.7	516	2 JE0134	mannan endo-1,4-beta
2	710	43.6	384	2 T37232	secreted beta-mann
3	692	42.5	363	2 S30386	mannan endo-1,4-beta
4	601	36.9	1331	2 A48954	mannan endo-1,4-beta
5	202	12.4	409	2 B25156	cellulase (EC 3.2.
6	200	12.3	488	2 A25156	cellulase (EC 3.2.
7	197	12.1	557	2 JC5487	cellulase (EC 3.2.
8	184.5	11.3	505	2 S39962	endoglucanase - Er
9	180	11.0	448	2 A27631	cellulase (EC 3.2.
10	176.5	10.8	466	2 C42360	cellulase (EC 3.2.
11	174.5	10.7	499	2 A27198	cellulase (EC 3.2.
12	174.5	10.7	504	2 S54784	cellulase (EC 3.2.
13	173.5	10.6	508	2 A26874	cellulase (EC 3.2.
14	162.5	10.0	486	2 I40548	bitfunctional cellu
15	162.5	10.0	508	2 G69593	cellulase (EC 3.2.
16	157	9.6	783	2 JC5467	cellulase (EC 3.2.
17	152	9.3	800	2 A29003	cellulase (EC 3.2.
18	152	9.3	822	2 JT0611	cellulase (EC 3.2.
19	151.5	9.3	499	2 JN0114	cellulase (EC 3.2.
20	151.5	9.3	825	2 JS0114	cellulase (EC 3.2.
21	149.5	9.2	357	2 PC4404	cellulase (EC 3.2.
22	149.5	9.2	941	2 S29043	cellulase (EC 3.2.
23	147.5	9.0	429	2 S29044	endoglucanase A pr
24	136.5	8.4	570	2 S56132	cellulase (EC 3.2.
25	131	8.0	814	1 CZCLDM	cellulase (EC 3.2.
26	130	8.0	426	2 A42649	cellulase (EC 3.2.
27	127	7.8	32	2 PC4278	guar gum-degrading
28	122.5	7.5	915	2 A43802	cellulase (EC 3.2.
29	122.5	7.5	1039	2 S02711	cellulase (EC 3.2.

30	122	7.5	438	2 A47702	glucan 1,3-beta-gl
31	117.5	7.2	428	2 S03767	cellulase (EC 3.2.
32	116	7.1	611	2 JC7177	endoglucanase V (E
33	115.5	7.1	747	2 B47093	cellulase (EC 3.2.
34	114	7.0	419	2 S72325	glucan 1,3-beta-gl
35	113.5	7.0	517	2 I40798	cellulase (EC 3.2.
36	111.5	6.8	441	2 A44815	cellulase (EC 3.2.
37	111	6.8	475	1 CZCLCA	cellulase (EC 3.2.
38	110	6.7	26	2 PC4285	guar gum-degrading
39	109.5	6.7	734	2 T14877	hypothetical prote
40	107	6.6	430	2 S55325	endo-beta-1,6-gluc
41	105	6.4	364	2 S12017	endoglucanase A -
42	105	6.4	456	2 T40276	probable gas1 fam1
43	104	6.4	556	2 A40589	cellulase (EC 3.2.
44	101.5	6.2	317	2 B72216	endoglucanase - Th
45	100.5	6.2	748	2 S19652	cellodextrinase C

ALIGNMENTS

RESULT 1
JE0134
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N: Alternate names: endo-1,4-beta-mannanase
C: Species: Bacillus circulans
C: Date: 03-Jun-1998 #sequence-revision 10-Jul-1998 #text-change 17-Mar-1999
C: Accession: JE0134
R: Yoshida, S.; Sako, Y.; Uchida, A.
Biosci. Biotechnol. Biochem. 62, 514-520, 1998
A: Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene c...
A: Reference number: JE0134; MUID:98233274
A: Accession: JE0134
A: Molecule type: mRNA
A: Residues: 1-516 <YOS>
A: Cross-references: DDBJ:AB007123
A: Note: the authors translated the codon CAA for residue 259 as Asn. CAA for residue
A: Note: the translation of the nucleotide sequence 1294-1365 is not given in this pap
C: Keywords: glycoprotein; glycosidase; hydrolase

Query Match	Score	Length	DB 2:	Length	DB 2:
Best Local Similarity	69.2%	Pred. No. 1e-83:			
Matches 207;	Conservative 46;	Mismatches 44;	Indels 0;	Gaps 0;	
QY	1	NANSGFYVSGTLVDANGNP	FMARGINHGAMWYRDOATTAEGIANTGANTVRIYLSDG	60	
DB	33	HAASGFYVSGTKLDANGOPFVARGVNAHWYDQSLTAIPATAKIGANTIRIVIANGH	92		
QY	61	QWTDDHTVNLISLAEDNHLVAVPEYHDATGSDIASLNRAVDYTEMRSALIGKEDT	120		
DB	93	KWTLDDVTVNNITLTCBQNTLIVLEVHDATGSDISDLNNAVWYIGIRKSALIGKEDR	152		
QY	121	VVINANWEGSWGDADGDKAIPRLRNAGNHTLVADVAAGGQPSGHIHYGREVF	180		
DB	153	VVINANWEGSWGDADGDKAIPRLRNAGNHTLVADVAAGGQPSGHIHYGREVF	212		
QY	181	NADPQNTFISIHMYEYAGNACVTRINIDRVNODLALVIGFGRHTNDVDDEATIMS	240		
DB	213	NADPQNTFISIHMYEYAGNACVTRINIDRVNODLALVIGFGRHTNDVDDEATIMS	272		
QY	241	YSEGRGVGLAMSWKGNKSPENEDLSDNDAAGNNLTAMGNTIVNGPYGLRHSRLSTV	299		
DB	273	YSNEGVGLAMSWKGNKSPENEDLSDNDAAGNNLTAMGNTIVNGPYGLRHSRLSTV	331		
RESULT 2					
T37232					
secreted beta-mannosidase - Streptomyces coelicolor					
C: Species: Streptomyces coelicolor					
C: Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 03-Dec-1999					
C: Accession: T37232					
R: Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rørdam, M.A.					


```

DB 234 DLAAD-----NPINDSNMYSVHF--YSGTNPISVDFNRDAMNSVRYALHGAAYFA 285
QY 222 GEFGRHRTNGDVDEATIMSYSEORGVGLAMSKNGKNGPEWEYIDLND----- 269
DB 286 TEMGTSLATG-----TTGPTL-AKADAWLDF-LNGNNISWNCFSISNKDEKAALNSLTS 338
QY 270 -----WAGNNLTAMGNTI---VNGPY 287
DB 339 LDPSGDKLMDNELTTSQGVYRARRKAGAY 367

RESULT 8
S39962
C:Species: Erwina carotovora
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39962
R:Cooper, V.J.C.; Salmond, G.P.C.
Mol. Gen. Genet. 241, 341-350, 1993
A:Title: Molecular analysis of the major cellulase (celV) of Erwina carotovora: evidenc
A:Reference number: S39962; MID:94067016
A:Accession: S39962
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-505 <COO>
A:Cross-References: EMBL:X76000; NID:g434941; PIDN:CAA53592.1; PID:g434942

```

```

Query Match 11.3%; Score 184.5; DB 2; Length 505;
Best Local Similarity 23.1%; Pred. No. 6.4e-07;
Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

```

```

QY 13 LVDANNPVPMGI-NHGAHYDQATTAEGTANGANTVRYLSDGCGWTD----- 66
DB 48 LVDEGKRVLQKGISSHGQWEGD-----YNNDSMKMLRDMGINY 89
QY 67 -----IHTVNLISLAEENHLVAVPEVHDATGYDSIASLRAYDYNT 108
DB 90 ERYAMTADAGTISNPSLANKVEVAAMASLQYIIIDPHILSDNDPRYKQAKTFRA 149
QY 109 EKRSLIGKEDVYIINANEWFG--SWEGD--AMADGYKALPRLNAGLNTIAMDAG 164
DB 150 EM-AGLYGSSPNYIEELANEPPNGVWNGQIRBYA--LEVTDTINSKDPDNLIIY---G 202
QY 165 WQGFPSIDHYGRVFNADPQRNTMSIHRTFAGSASVRYNRINDRYLQDLALVIGER 224
DB 203 TGTWSDIDH--ADNQLPDPPTMTALHP--YAGTHGQFLRDRIDYASRGAIIVSEV 257
QY 225 GHRHTNGD-----VDEATIMSYSEORGVGLAMSKNGKNGPEWEYL-----DLSNDWAGNNL 275
DB 258 GTSDSAGNGPFLPESQWTWIDFLNNGVSWVWNSLIDKSEASALAPGASKSGGWTEDNL 317
QY 276 TAMGNTI 282
DB 318 STSGKEV 324

```

```

RESULT 9
A27631
C:Function: cellulase (EC 3.2.1.4) precursor - Clostridium acetobutylicum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium acetobutylicum
C:Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
C:Accession: A27631
R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.
Appl. Environ. Microbiol. 54, 1269-1292, 1988
A:Title: Structure of an endo-beta-1,4-glucanase gene from Clostridium acetobutylicum P2
A:Reference number: A27631; MID:88268074
A:Accession: A27631
A:Molecule type: DNA
A:Residues: 1-448 <ZAP>
A:Cross-References: EMBL:M31311; NID:g144789; PIDN:AAA23230.1; PID:g144790
A:Note: the authors translated the codon GAG for residue 116 as Gly, GAA for residue 263

```

```

C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

```

```

Query Match 11.0%; Score 180; DB 2; Length 448;
Best Local Similarity 22.1%; Pred. No. 1.2e-06;
Matches 71; Conservative 59; Mismatches 135; Indels 56; Gaps 14;

```

```

QY 8 VSGTLLYDANGNPFVNRGI-NHGAHY-----KQDQATTAIEGIANTGANTVRI 54
DB 49 VVGSQGLCDNSGNKPIQKSGSHGLQVYFVYDSMKELRDK-----WGVNVA 98
QY 55 VL--SDGCGWT--KDDIHTVNLISLAEENHLVAVPEVHDATGYDSIASLRAYDIWIM 110
DB 99 AMTNEGGSYISNPSQKREKIRIVQDAIDLNNYIIDPHILSDNDNPNRYEQAQSEFQEM 158
QY 111 RSALIGKEDVYIINANEWFGSMEGDADGYK-----QAIPLRNAGLNTIAMDAGWG 166
DB 159 AEE-YKYSNVYIEICNEENG--GTMMANDIKPYANYIIPALRAIDPNNIITVGTSTWS 214
QY 167 QEPQSIDHYGRVFNADPQR--NTMFSIHMYEYAGNASQVRYNRINDRYLQDLALVIGER 224
DB 215 QDVDDIAD-----NPLRGINMTYCHF--YAGTHQSLDKINYASKGAIIVTEM 264
QY 225 GHRHTNGD-----VDEA-TIMSYSEORGVGLAMSKNGKNGPEWEYL-----DLSNDWAGNNL 275
DB 265 GTSDSAGNGPFLYDESOQWVDEMASKNISMTWMLALCDKSEASALKSGSSTGGWTDSDL 324
QY 276 TAMGNTIYGPGLRSTSL 296
DB 325 TTSGLFVKKSGISGNTTSOTS 345

```

```

RESULT 10
CA2360
C:Function: cellulase (EC 3.2.1.4) E5 precursor - Thermomonospora fusca
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Thermomonospora fusca
C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 20-Mar-1998
C:Accession: CA2360
R:iao, G.; Ghangas, G.S.; Jung, E.D.; Wilson, D.B.
J. Bacteriol. 173, 3397-3407, 1991
A:Title: DNA sequences of three beta-1,4-endoglucanase genes from Thermomonospora fus
A:Reference number: A42360; MID:91258320
A:Accession: CA2360
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-466 <LAD>
A:Cross-References: GB:L01577; NID:g154693; PID:g154694
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulase-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:38-137/Domain: bacterial cellulase-binding domain homology <BCB>

```

```

Query Match 10.8%; Score 176.5; DB 2; Length 466;
Best Local Similarity 25.0%; Pred. No. 2.4e-06;
Matches 78; Conservative 48; Mismatches 125; Indels 61; Gaps 17;

```

```

QY 8 VSGTLLYDANGNPFVNRGIN-HGAHYKQDQAT-TAIEGIA-NTGANTVRI--VLSGCGW 62
DB 173 VCGTQCDSEHGNPVOLRGKSTGIQWFDHCLDSSLDALAYDMKADIRLTSYIGEDGYE 232
QY 63 TDDDIHTVA--NLISLAEENHLVAVPEVHDATGYDSIASLRAYDIWIMSKALLGKEPT 120
DB 233 TNPREGTIDIDILIMATARGLYVIVDWHLPLPGDPHYVLDRAKTFEAFIAQRHASKTN- 291
QY 121 VTIINANEWFGSMEGDAMD--GY-KQAIPLRNAGLNTIAMDAGWGQFQSGIHDYGR 177

```

Db 292 VLYEIANE-----PNCVSASIKSYAEVYIPVIRORDPDSPVITVGRGSSLSGV-EGSGP 346
QY 178 EVFNADP--QRTNFSIMYETAGNAGNOVNTNIDRVINODLAL--VIGEGHRTNGD 232
Db 347 AEIAPNPNASINIMAFHFY-----AASHRDNYINALREASELPVYVTEGETETYG 400
QY 233 -----VDENTIMSISQROGVGLAWSK-----GNGPEWEYLDLSNDW 270
Db 401 GANDFQMDR--YIDLMAERIKIGTWKMYSDDFRSGAVFOPCTCASGP-----W 448
QY 271 AGNNLTANGNTI 282
Db 449 SSSSLKASGQWY 460

RESULT 11
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 15-Dec-1988 #sequence, revision 15-Dec-1988 #text, change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Dozum, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397
A:Molecule type: DNA
A:Accession: A27198
A:Residues: 1-499 <NNA>
A:Cross-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671
A:Experimental source: strain IFO3034
A:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1.36/Domain: signal sequence #status predicted <SIG>

Query Match 10.7%; Score 174.5; DB 2; Length 499;
Best Local Similarity 21.1%; Pred. No. 3.8e-06;
Matches 70; Conservative 66; Mismatches 123; Indels 73; Gaps 15;

QY 3 NSGFYVSTIYDANGNFVNRGT-NHGHNATKQDATTAGIANTGANTYRIYSDGGQ 61
Db 39 NGQISIKGTQLVNRDGNKAVOLKISSRGLQWYGD-----FVKKDSIK 80
QY 62 WTKDD-----IHT-----VRNLISLAEDNHLVAVPEVHDATGYDSIA 98
Db 81 WLRDMDGTVFRAMATYDGGYIDNPVKYKKEVNAKEIGIYIITDHLNDGNFQ 140
QY 99 SLNRADVIMRSALIGKEDYIINIANEFG--SWEGD--AMADGYKQAIPLRNAGL 154
Db 141 HKEKAKFFKEM--SLYGNTPNVIYEIANEPNGDVNFKRDIKPYAE--EYLSVIRKNDP 196
QY 155 NHTLNVDAAGGQFPOSIDHGREVFNDPQ--RNTFESIMYETAGNAGNOVNTNIDR 211
Db 197 DNIITV--GIGTWSQVDN-----AADQDLKQANVYALHF--YAGTHGQSLDCKRNY 245
QY 212 VLNODLALVIEGFGRHRTNGD--VDEA-TIMSYSEORGVGLAWSKNGPEWEY-- 264
Db 246 ALISGAPLIFYEWGTSDSAGNGVPLDQSRKMLNTLDSKNTISWVWNLSDKOESSALKP 305
QY 265 --DISNDWAGNNLTANGNTIYNGPYGLRETSR 294
Db 306 GASRTGGMPLTDLTASGTFVRENILGNRDSRK 337

RESULT 12
S54744
cellulase (EC 3.2.1.4) CelV1 precursor - Erwinia carotovora (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Erwinia carotovora
A:Variety: SCC 3193

C>Date: 27-Oct-1995 #sequence, revision 03-Nov-1995 #text, change 08-Oct-1999
C:Accession: S54744; S44996
R:Mae, A.; Heikkinen, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A:Title: Structure and regulation of the Erwinia carotovora subspecies carotovora SCC
A:Reference number: S54744; MUID:95231512
A:Accession: S54744
A:Molecule type: DNA
A:Residues: 1-504 <MAE>
A:Cross-references: EMBL:X79241; NID:g493492; PIDN:CAA55823.1; PID:g493493
C:Genetics:
A:Gene: celV1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1.32/Domain: signal sequence #status predicted <SIG>
F:3.3-504/Product: cellulase #status predicted <MAT>

Query Match 10.7%; Score 174.5; DB 2; Length 504;
Best Local Similarity 23.7%; Pred. No. 3.9e-06;
Matches 73; Conservative 43; Mismatches 123; Indels 69; Gaps 14;

QY 13 LYDANGNFFVNRGT-NHGHNATKQDATTAGIANTGANTYRIYSDGGQTKDD--IHT 69
Db 48 LVDEGKRVOLRGISSNGLQWYGD-----YVKKDSMKWLRDPMGJNV 89
QY 70 VRNLISLAEDNHLVAVPEV-----HDATGYDSIASLNRADVW 107
Db 90 FRVAVYTAE--NGYINPBLANKVEAVAAOGLGYIITDHLTSDNDPNTYKQAKTFE 148
QY 108 IEMRSALIGKEDYIINIANEFGS--WEGD--AMADGYKQAIPLRNAGNHTLNVDA 163
Db 149 AEM-AGLIGSNPNVYIEIANEPNGSVTNNGIIRPYA--LEVTDTRIRKDDNLIIVSG 204
QY 164 GNGQFPOSIDHGREVFNDPQRTNFSIMYETAGNAGNOVNTNIDRVINODLALVIGE 223
Db 205 TWS--QDIHD--AADQLPDPNTLYALHF--YAGTHGQFLRDIRDYAOSRGAIFVSE 256
QY 224 FGHRTNGD-----VDENTIMSISQROGVGLAWSKNGPEWEY--DLSNDWAGNN 274
Db 257 WGTSDASNGGPFPLPESQTIWIDFLNNGISWVWNLSDKSETSAALVAGASKSGGWTQCN 316
QY 275 LTAMGNTI 282
Db 317 LSTSGKFV 324

RESULT 13
A26874
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 08-Mar-1989 #sequence, revision 08-Mar-1989 #text, change 15-Oct-1999
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.H.
J. Bacteriol. 169, 2017-2025, 1987
A:Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.
A:Reference number: A26874; MUID:87194581
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <ROB1>
A:Cross-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <ROB2>
A:Experimental source: strain DLG
A:Note: The authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation
C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; polys
F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIC>

Query Match	10.68	Score 173.5	DB 2	Length 508
Best Local Similarity	20.88	Pred. No. 4.7e+06		
Matches 69	Conservative 68	Mismatches 122	Indels 73	Gaps 15

[illegible]

RESULT 14
I40548
bifunctional cellulase precursor - *Bacillus* sp.
C:Species: *Bacillus* sp.
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40548
R:Han, S.J.; Yoo, Y.J.; Kang, H.S.
J. Biol. Chem. 270, 26012-26019, 1995
J>Title: Characterization of a bifunctional cellulase and its structural gene: the cel
A:Reference number: I40548; MUID:96029707
A:Accession: I40548
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-486 <RES>
A:Cross-references: EMBL:U27084; NID:g857575; PIDD:AA643478.1; PID:g857576
C:Genetics:
A:Gene: cel

[illegible][illegible]

RESULT 15
669593
cellulase (EC 3.2.1.4) bglc precursor - *Bacillus subtilis*
N:Alternate names: endo-1,4-beta-glucanase
""""""""""

C:Date: 05-Dec-1997
C:Accession: 669593
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bar
A.; Brown, S.; Broilletter, S.; Brusch, C.V.; Caldwell, B.; Capuano, J.; Carter, N.M.;
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funae, S.; Galizzi, A.; Gal
lech, U.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koester, P.; Konigsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lepidus, A.; Lindo
A:Authors: Lauber, U.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y. M.; Ogawa, K.; Ogizawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Poter
Rieger, M.; Rivolta, C.; Roch, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scall
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekovska, A.; Se
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekovska, A.; Se
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekovska, A.; Se
T.; Winters, P.; Wipst, A.; Yanaka, T.; Terstets, P.; Tomono, A.; Tosato, V.; Uchida
A:Authors: Yoshikawa, H. F.; Zimmstein, E.; Yoshikawa, H.; Yano, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
A:Reference numbers: A69580; MUID: 98044033

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <KIN>
A:Cross-references: GB:299113; GB:AL009126; NID:g2534090; PIDN:CAAB3696.1; PID:e11114
A:Experimental source: strain 168
A:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, P.
Nucleic Acids Res. 14, 9159-9170, 1986
A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
A:Reference number: A26114; MUID:87065783
A:Accession: A26114
A:Molecule type: DNA
A:Residues: 10-508 <MAC>
A:Note: part of this sequence, including the amino end of the mature form, was confirmed by R.Lindahl, V.; Aa, K.; Tronsmo, A.
Antonie Van Leeuwenhoek 66, 327-332, 1994
A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis
A:Reference number: I40353; MUID:95225656
A:Accession: I40353
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN1>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
R.Lindahl, V.; Aa, K.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24239
A:Accession: S24239
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN2>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
R.Wolf, M.; Geczi, A.; Borris, R.
submitted to the EMBL Data Library, December 1993
A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: C
A:Reference number: S49103
A:Accession: S49103
A:Molecule type: DNA
A:Residues: 10-508 <HOL>
A:Cross-references: EMBL:Z29076; NID:g509266; PIDN:CAA82317.1; PID:g509267
R.Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:44:15 ; Search time 131.22 Seconds
(without alignments)
73.041 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGFVSTTLYDANGNP.....TIVNGPYGLRSTSLSTVFT 300

Scoring table:
BIOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues
Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	692	42.5	383	1	MANA_STRLI
2	601	36.9	1331	1	MANB_CALSA
3	202	12.4	409	1	GNB2_BACSA
4	201	12.3	400	1	GNB5_BACAG
5	200	12.3	488	1	GNB1_BACSA
6	184.5	11.3	505	1	GNB4_BACSA
7	180.5	11.1	444	1	GNB5_ERMCA
8	180	11.0	448	1	GNB5_ERMCA
9	177.5	10.9	466	1	GNB5_ERMCA
10	174.5	10.7	504	1	GNB5_ERMCA
11	173.5	10.6	499	1	GNB1_BACSA
12	162.5	10.0	499	1	GNB2_BACSA
13	161	9.9	459	1	GNB1_BACSA
14	152	9.3	800	1	GNB1_BACSA
15	151.5	9.3	499	1	GNB3_BACSA
16	151.5	9.3	825	1	GNB3_BACSA
17	149.5	9.2	941	1	GNB1_BACSA
18	147.5	9.0	429	1	GNB1_BACSA
19	142	8.7	562	1	GNB1_BACSA
20	141	8.7	312	1	GNB4_BACSA
21	131	8.0	814	1	GNB4_BACSA
22	130	8.0	426	1	GNB4_BACSA
23	122.5	7.5	1039	1	GNB4_BACSA
24	122	7.5	438	1	GNB4_BACSA
25	117	7.2	426	1	GNB4_BACSA
26	115.5	7.1	747	1	GNB4_BACSA
27	113.5	7.0	517	1	GNB4_BACSA
28	111	6.8	475	1	GNB4_BACSA
29	109.5	6.7	440	1	GNB4_BACSA
30	105	6.4	364	1	GNB4_BACSA
31	104	6.4	566	1	GNB4_BACSA
32	100.5	6.2	748	1	GNB4_BACSA
33	99.5	6.1	448	1	GNB4_BACSA

34	98	6.0	584	1	GNB4_BACSA	P25472 clostridium
35	97.5	6.0	375	1	GNB4_BACSA	P23044 robiliarda
36	97.5	6.0	647	1	GNB4_BACSA	P08331 escherichia
37	97	6.0	493	1	GNB4_BACSA	P19487 xanthomonas
38	95	5.8	515	1	GNB4_BACSA	P28623 clostridium
39	94.5	5.8	645	1	GNB4_BACSA	P40601 photorhabd
40	94.5	5.8	856	1	GNB4_BACSA	P37898 saccharomyc
41	94	5.8	562	1	GNB4_BACSA	P52911 saccharomyc
42	94	5.8	890	1	GNB4_BACSA	P08696 clostridium
43	93.5	5.7	406	1	GNB4_BACSA	P16216 rumiincoccu
44	93.5	5.7	510	1	GNB4_BACSA	O13692 schizosacch
45	93.5	5.7	566	1	GNB4_BACSA	P23550 paenibacilli

ALIGNMENTS

RESULT 1	MANA_STRLI	STANDARD:	PRT:	383 AA.
AC	P51529:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	MANNA ENDO-1.4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-MANNANASE) (1.4-BETA-D-MANNAN MANNANOHYDROLASE).			
GN	MANA.			
OS	Streptomyces lividans.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
CC	Actinomycetales; Streptomyces; Streptomyces; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 36-42.			
RC	STRAIN-66 / 1326;			
RA	MEDLINE: 93207541.			
RA	Arcand N., Kluepfel D., Paradis F.W., Morosoli R., Shareck F.;			
RT	"Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence			
RL	Biochem. J. 290:657-663(1993).			
RN	[2]			
RP	REVISIONS TO C-TERMINUS.			
RC	STRAIN-66 / 1326;			
RA	Shareck F.;			
RL	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES			
CC	CELSIUS AND A PH OF 6.8.			
CC	-1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-			
CC	LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND			
CC	GALACTOGLUCOMANNANS.			
CC	-1- SUBUNIT: MONOMER.			
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: M92297; AAA26710.2; -.			
DR	INTERPRO: IPR001547; -.			
DR	PFAM: PF00150; cellulase; 1.			
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Hydrolase; Glycosidase; Signal.			
FT	SIGNAL			
FT	CHAIN 36 383			
FT	DOMAIN 336 340			
FT	SEQUENCE 383 AA; 39681 MW; SDB4B407C64E94C3 CRC64;			
SO	MANNA ENDO-1.4-BETA-MANNOSIDASE.			

Query Match 42.5%; Score 692; DB 1; Length 383;
Best Local Similarity 46.7%; Pred. NO. 2.9e-46;

	MATCHES	141:	Conservative	51:	Mismatches	104;	Indels	6;	Gaps	5;
OY	2	ANSQGYVSGTLLIYDANGPFPVRGGINHGHANTKQDAITAIAGCIANTGANTRYIVLSOGGQ	61							
Dd	36	AAGGJHVSNGRVEBENGSAFVRGNAYTYTPDR-TSSIDADIARKGANTRYAVLSSGG	94							
OY	62	WTKDIDHYRNLIStAEENHLVAPEVDADGY--DSIASINRAVDYIEKRSAL-IGKE	118							
Dd	95	WTKTSASVSLIIIGCKANKXYICYLEVEDDTYGIGDGATSLDDAGDYGVGRKSAMRRQE	154							
OY	119	DTVITIANEMFGSWEGDAMADGYKOAPRLNACLNHTLTAAADAAGMGQ-FPOSIDHYGR	177							
Dd	155	DYVVVNINNEPEGNNNTYAAMDATSTSAIGKLKGALMADPAVMOCODMSGTIRSNAA	214							
OY	178	EVENNDPCQRNFESTIHMEEYAGNSOVRTNDDTVLNDGLAIVEGFGRHNGDPVDEAT	237							
Dd	215	SVFASDPDRNYTFSTHMHG-VTDTALEARDYTINALVGNGLPYVGEFFGDSDGPNPDDA	273							
OY	238	IMSYSRGQVGTLANSMKGNGBEWETLDLSNDMAANNLTAMGNTITVNGPYGLRETSRLST	297							
Dd	274	IMATAQSIGVGYLGWSWGNGGGVEYELDMVNGFDPSNLSTLMGCRILLYSGNGLAATSRTAT	333							
OY	298	VF 299								
Dd	334	VF 335								
	RESULT	2								
	MANB_CALSA	STANDARD:	PRT: 1331 AA.							
ID	MANB_CALSA	STANDARD:	PRT: 1331 AA.							
AC	P22533:									
DT	01-AUG-1991 (Rel. 19, Created)									
DT	01-JUL-1993 (Rel. 26, Last sequence update)									
DT	15-JUL-1999 (Rel. 38, Last annotation update)									
DE	BETA-MANNAOSE/ENDOGLUCANASE A PRECURSOR [INCLUDES: MANNAN ENDO-1,4-BETA-MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNAOSE) (ENDO-1,4-MANNAOSE); ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULOSE)].									
GN	MANA.									
OS	Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).									
OC	Bacteria; Firmicutes; Bacilli; Clostridium group;									
OC	Thermoanaerobacter group; Caldicellulosiruptor.									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RX	MEDLINE: 93119139.									
RA	GIBBS M.D., Saul D.J., Luthi E., Bergquist P.L.;									
RT	"The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme."									
RL	Appl. Environ. Microbiol. 58:3864-3867(1992).									
RP	[2]									
RP	SEQUENCE OF 1-346 FROM N.A.									
RX	MEDLINE: 91247819.									
RA	Luehl E., Jasmat N.B., Greyling R.A., Love D.R., Bergquist P.L.;									
RT	"Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'."									
RL	Appl. Environ. Microbiol. 57:694-700(1991).									
CC	-1- FUNCTION: DEBRASSATION OF HEMI CELLULOSES. THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOGLUCANASE ACTIVITIES.									
CC	-1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS.									
CC	-1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.									
CC	-1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES CELSIUS.									
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).									
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO CELLULOSE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).									
CC	-----									
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -									

```
CC The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/ebis/ or send an email to license@ebi.ac.uk).
```

```
CC -----
```

```
DR EMBL; L01257; AAA71887.1; -
```

```
DR EMBL; M36063; AAA72861.1; -
```

```
DR PIR; B43745; B43745.
```

```
DR PIR; A48954; A48954.
```

```
DR HSSP; Q06851; 1NBC.
```

```
DR INTERPRO; IPR001547; -
```

```
DR INTERPRO; IPR001956; -
```

```
DR PFAM; PF00942; CBD_3; 2.
```

```
DR PFAM; PF00150; cellulase; 1.
```

```
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
```

```
KW Hydrolyase, Glycosidase, Cellulose degradation, Signal;
```

```
KM Multifunctional enzyme.
```

```
FT SIGNAL 1 41 POTENTIAL.
```

```
FT CHAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.
```

```
FT DOMAIN 42 325 CATALYTIC (MANNANASE ACTIVITY).
```

```
FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).
```

```
FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).
```

```
FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).
```

```
FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).
```

```
FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).
```

```
FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).
```

```
FT ACCT_SITE 162 162 PROTON DONOR (BY SIMILARITY).
```

```
FT ACCT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
```

```
FT CONFLICT 338 338 T->P (IN REF. 2).
```

```
FT CONFLICT 340 346 TPRTPTPT->ROHQHQ (IN REF. 2).
```

```
SQ SEQUENCE 1331 AA; 146892 MM; FFBGCA5IBB8DF0E0 CRC64;
```

```
Query Match 36.9%; Score 601; DB 1; Length 1331;
```

```
Best Local Similarity 44.5%; Pred. No. 1,4e-38;
```

```
Matches 125; Conservative 42; Mismatches 108; Indels 6; Gaps 4;
```

```
OY 25 GINGHANTKDOATTAISGIANTGANTRYRIYVSDSGOWTKDIDHYRNRLISAEDNHLYA 84  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 55 GTNHACHYRDLPLDALNGIRSMGNNSRVVYSNYRWKRPBASEVAANISLSRSLGFRA 114  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
OY 85 -VEPHDATGY---DSIASLRADVDTWEEMRSALIGKEDETVIINIANEFGSMEGDAMD 140  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 115 ILLEPHDTTIGSEDAAGCSLAQAVETKAEIKSVLDGNEDEFVIINIGNEYGNANNQNVYN 174  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
OY 141 GTKQAPRLRNGLNHTLVADAAGNQ-FPGSIDHYGREVENADPORNTMESIHWTAYAG 199  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 175 DTRNAIKALFDGFKHTIIWVDAPNNNGOMSNMTNRNASIMADPLRNIVFSIHMYG-VY 233  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
OY 200 GNASGVRRNIIDVLMODLALVTGEGRHRTNDVDVDEATIMSYSEORGVMIAWSMGNGP 259  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 234 NTASTVEERYIKSFVKNGLFVLIGEFGHQITDDPDDEALVRAKQKYIGLFPSWCGNSS 293  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
OY 260 EWEYILDLSNDVANMLTAGNTIVNGPYGLARETSRLSTVET 300  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 294 YVGVIDMVNMDPNNPTPMGWGYKTNAIGTSTPTPTSTVET 334  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
RESULT 3  
GUN2_BACS4 STANDARD; PRT; 409 AA.  
AC GUN2_BACS4 P06565;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-JUL-1993 (Rel. 38, Last annotation update)  
DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-L,4-BETA-GLUCANASE) (CELLULOSE).  
GN CELB.  
OS Bacillus sp. (strain N-4).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.
```

RX MEDLINE: 87056924.
 RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RC *Bacillus* sp. strain N-4 and their strong homology.";
 RL J. Bacteriol. 168:479-485(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOLACTOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14729; AAA22299.1; -
 DR PIR: B25156; B25156.
 DR INTERPRO: IPR001547; -
 DR PIRAM: PIR00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase.
 FT ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 409 AA; 45690 MW; 98866066B4DA5F CRC64;

 Query Match 12.4%; Score 202; DB 1; Length 409;
 Best Local Similarity 25.0%; Pred. No. 1.6e-08;
 Matches 70; Conservative 46; Mismatches 98; Indels 64; Gaps 14;

 QY 8 VSGTLLDANGNPFVNGI-NHGAWYKQATTAEGTANTGATVRLVLSGGQWTKDD 66
 DB 40 ISNGELVNRGEPOVLKGMSSHGLQWTC-----GQVNESKMKLDD 81
 QY 67 -----IHT-----VANLSLAEDNHLVAPVEYDAGTGSIALSLNRA 103
 DB 82 WGTTFRAAMTSSGGYIEDPSYKEKEVAEALDGIYIIDWHILSDNDPNYKEEA 141
 QY 104 VDYWIERSALFQKEDVYIINIANEMFGS--WEGD--AMADGYKQALPRLNAGLNTL 158
 DB 142 KDFEDEM-SELXGDPNVIYIETIANEPNSDVTWNOIPIYAE--EVLPIRANNDPNNTI 197
 QY 159 WDAAGNGQFQSIHDYGREVFADPQRTMFSIHMEYAGNAGNSOVRTNIDRYLNDPLA 218
 DB 198 IV---GRTGMSQDVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLADQVDYALDQGA 249
 QY 219 LVIGEFGRHRTNGD---VDEATL-MSYSEORGVGLWMS 253
 DB 250 IFVSEWGTSEATGDGVFLDEAQVWIDFMDERNLISWANS 289

 RESULT 4
 GUN5_BACAG STANDARD; PRT; 400 AA.
 AC 085465;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ENDOGLUCANASE 5A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (ALKALINE
 DE CELLULASE).
 GN CEL5A.
 OS *Bacillus agaradhaerens* (*Bacillus agaradhaerens*).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC *Bacillus*/Staphylococcus group; *Bacillus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 8721;
 RA Bjornvad M.E.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329.
 RC STRAIN-AC13 / NCIMB 40482;
 RX MEDLINE: 98153671.
 RA Davies G.J., Dauter M., Brzozowski A.M., Bjornvad M.E.,
 RA Andersen K.V., Schuelein M.;
 RT "Structure of the *Bacillus agaradhaerens* family 5 endoglucanase at 1.6-
 RT A and its cellobiose complex at 2.0-A resolution.";
 RL Biochemistry 37:1926-1932(1998).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329.
 RC STRAIN-AC13 / NCIMB 40482;
 RX MEDLINE: 98384136.
 RA Davies G.J., MacKenzie L.F., Varrot A., Dauter M., Brzozowski A.M.,
 RA Schuelein M., Withers S.G.;
 RT "Snapshots along an enzymatic reaction coordinate: analysis of a
 RT retaining beta-glucoside hydrolase.";
 RL Biochemistry 37:11707-11713(1998).
 CC -1- CATALYTIC ACTIVITY: ENDOLACTOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF067428; AAC19169.1; -
 DR PDB: 1A3H; 16-MAR-99.
 DR PDB: 2A3H; 16-MAR-99.
 DR PDB: 3A3H; 16-MAR-99.
 DR INTERPRO: IPR001547; -
 DR PIRAM: PIR00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
 FT ACT_SITE 165 165 PROTON DONOR.
 FT ACT_SITE 254 254 NUCLEOPHILE.
 SQ SEQUENCE 400 AA; 44702 MW; 3F9C66FB9BC36FF CRC64;

 Query Match 12.3%; Score 201; DB 1; Length 400;
 Best Local Similarity 26.6%; Pred. No. 1.8e-08;
 Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

 QY 8 VSGTLLDANGNPFVNGI-NHGAWY-----KQATTAEGTANTGATVRI 54
 DB 40 ISNGELVNRGEPOVLKGMSSHGLQWYGFVNESKMKLR-----DWGINFRA 89
 QY 55 VL--SDGQWTKDI-HTVRLISLAEDNHLVAPVEYDAGTGSIALNRAVYWIEMR 111
 DB 90 AMYTSSSGGYIDPSYKEKEKAVAEALDGIYIIDWHILSDNDPNYKEAKPFEDM- 148
 QY 112 SALIGKEDVYIINIANEMFGSWECDANADGYK---QALPRLNAGLNTLMDVDAAGNQ 167
 DB 149 SELXGDPNVIYIETIANEPNS--DVTWNOIKPYAEVILPIRANNDPNNTIIV---GRT 203
 QY 168 FQSIHDYGREVFADPQRTMFSIHMEYAGNAGNSOVRTNIDRYLNDLALVIGEFGR 227
 DB 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLADQVDYALDQGAIFVSEWGT 258
 QY 228 HTNGD---VDEATL-MSYSEORGVGLWMS 253
 DB 259 AATGDGVTFLDEAQVWIDFMDERNLISWANS 289

 RESULT 5
 GUN1_BACS4 STANDARD; PRT; 488 AA.
 AC P06566;
 ID GUN1_BACS4

01-JAN-1988 (Rel. 06, Created)
 01-JAN-1988 (Rel. 06, Last sequence update)
 15-JUL-1999 (Rel. 38, Last annotation update)
 ENDOGUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
 CELA.
 OS Bacillus sp. (strain N-4).
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87056924.
 RA Fukumori F., Sasahara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RT Bacillus sp. strain N-4 and their strong homology";
 RL J. Bacteriol. 168:479-483(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14781; AAA2301.1; -
 DR PIR: A25156; A25156.
 DR INTERPRO: IPR001547; -
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE, PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE, PS00659; Hydrolase; Glycosidase.
 KW Cellulose degradation; Hydrolase; Proton donor (by similarity).
 FT ACT_SITE 163 163
 FT ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 488 AA; 54264 MW; 97248BA53D28D3B5 CRC64;
 SO SEQUENCE

Query Match 12.3%; Score 200; DB 1; Length 488;
 Best Local Similarity 24.6%; Pred. No. 2.8e-08;
 Matches 70; Conservative 49; Mismatches 103; Indels 62; Gaps 14;

01 3 NSGFVSTTLYDANGNPFVNRGI-NHGHWKQDQATTAIESIANTGANTVRIYSDGGQ 61
 DB 33 NGQLSIQNGQLVNEHGDVQLKMSHGLQWYG-----QFVNTDSIK 74
 QY 62 WTQKD-----IHT-----VRNLISLAEDNHLVAPEVDATGYDSTA 98
 DB 75 WLRDWMGITVFRAMYTSSGYIEDPSVKEKVEAIDIGIYIIDMHLSDNDPNI 134
 QY 99 SLNRADVWIERKSLALIGKEDVIYINIANEWFG---SWEGDAMADGY-KQALPRLNAGL 154
 DB 135 YKEEKKEFFDEN-SALYGYIPVNYIIEIANEPGNHVRW--DSHIRPYAEVPIVRANDP 191
 QY 155 NHTLWADAGMGQFQSIHDYGREYFNADPQRTMFSIMHEYAGNAGNSQVETNIDRYLN 214
 DB 132 NNIVIVGRTWV---QDVHEADNOLD-DP--NVMYAFHF--YAGHGGQLKNQVDYALS 243
 QY 215 ODALVIGFEGHRTNGD---VDEATI-MSYSEORGVGWILMS 253
 DB 244 RQALIFVSEMTSATGDCGVFLDEKQVWIDEMDERNLISWAMNS 287

RESULT 6
 GUNV_ERMCA STANDARD: PRI; 505 AA.
 AC 047096;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ENDOGUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)
 DE (CELLULOSE V).

GN CELV.
 OS Erwilia carotovora.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Pectobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-SCRI193;
 RX MEDLINE: 94067016.
 RA Cooper V.J.C., Salmond G.P.C.;
 RT "Molecular analysis of the major cellulase (celv) of Erwilia
 RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
 RT domains";
 RL Mol. Gen. Genet. 241:341-350(1993).
 CC -1- FUNCTION: ENDOGUCANASE WITH SOME EXOGUCANASE ACTIVITY. THE PH
 CC OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
 CC CELSIUS.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X76000; CAA53592.1; -
 DR HSSP: Q06851; INBC.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001956; -
 DR PFAM: PF00942; CBD_3; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE, PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE, PS00659; Hydrolase; Glycosidase; Signal.
 KW CELLULOSE degradation; Hydrolase; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 505
 FT DOMAIN 32 334
 FT DOMAIN 335 352
 FT DOMAIN 353 505
 FT ACT_SITE 168 168
 FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 505 AA; 54900 MW; DBEA9337BBAD2623 CRC64;
 SO SEQUENCE

Query Match 11.3%; Score 184.5; DB 1; Length 505;
 Best Local Similarity 23.1%; Pred. No. 4.6e-07;
 Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

01 13 LYDANGNPFVNRGI-NHGHWKQDQATTAIESIANTGANTVRIYLSGGQVTKND----- 66
 DB 48 LVDEQGRVQLRGISSHGLQWFGD-----YNNKSMKWLNRDWMGINV 89
 QY 67 -----IHTVRNLISLAEDNHLVAPEVDATGYDSTANRAVDYWI 108
 DB 90 FRVAMYTAAADGYISNPSLANVKEVAAQSLGYIIIDMHLSDNDNPIYKQAKTFEA 149
 QY 109 EKRSLALIGKEDVIYINIANEWFG--SWEGD--AWADGYKQALPRLNAGHNTLWADAG 164
 DB 150 EK-AGLYGSPNVYIIEIANEPGNHVRW--DSHIRPYAEVPIVRANDP 202
 QY 165 MGQFQSIHDYGREYFNADPQRTMFSIMHEYAGNAGNSQVETNIDRYLNODLALVIGEF 224
 DB 203 TGTWSQDHDH--AADNOLPDPNTWYALHF--YAGTHQFPLRDRDYAQSQAALFVSEW 257
 QY 225 GHRHTNGD---VDEATIMSSEORGVGWILMSKNGPMEWYL---DISNDWAGNLL 275
 DB 258 GTSDSAGSGGFLPLPSQWIDFNNRGVSWNWSLITDSEASALAPASNSGGTTEQNL 317
 QY 276 TAWGNTI 282

```

RESULT      7
ID           GUNN_ERMCA      STANDARD:      PRT;      444 AA.
AC           Q59394;
DT           01-NOV-1997 (Rel. 35, Last created)
DT           01-NOV-1997 (Rel. 35, Last sequence update)
DT           01-NOV-1997 (Rel. 35, Last annotation update)
DE           ENOGLUCANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE N)
GN           CELN.
OS           Erwania carotovora.
OC           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC           Pectobacterium.
RN           [1]
RP           SEQUENCE FROM N.A.
RC           STRAIN-ATROSEPTICA FCBR C18;
RX           MEDLINE: 98299944.
RA           Olsen O., Thomsen K.K., Weber J., Dues J.O., Svendsen I., Wegener C.,
RA           von Wettstein D.;
RT           "Transplanting two unique beta-glucanase catalytic activities into
RT           one multienzyme, which forms glucose.";
RL           Biotechnology 14:71-76(1996).
CC           -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC           LINKAGES IN CELLULOSE.
CC           -1- SUBCELLULAR LOCATION: SECRETED.
CC           -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC           HYDROLASES).
-----
CC           This SWISS-PROT entry is copyright. It is produced through a collaboration
CC           between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC           the European Bioinformatics Institute. There are no restrictions on its
CC           use by non-profit institutions as long as its content is in no way
CC           modified and this statement is not removed. Usage by and for commercial
CC           entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC           or send an email to license@isb-sib.ch).
-----
EMBL: L39788; AAC37033.1; -.
DR           INTERPRO: IPR001547; -.
DR           INTERPRO: IPR001956; -.
DR           PFMW; PF00942; CBD_3; 1.
DR           PFMW; PF00150; cellulase; 1.
DR           PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KM           Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT           SIGNAL
FT           1
FT           CHAIN
FT           32
FT           ACT_SITE
FT           168
FT           ACT_SITE
FT           256
FT           SEQUENCE
FT           444 AA; 48300 MW; FA7E4179004CBB43 CMC64;
FT           PROTON DONOR (BY SIMILARITY).
FT           NUCLEOPHILE (BY SIMILARITY).
-----
Query Match      11.1%; Score 180.5; DB 1; Length 444;
Best Local Similarity 22.5%; Pred. No. 7.9e-07;
Matches 69; Conservative 48; Mismatches 123; Indels 67; Gaps 12;

13  LYDANGNFVWRCI-NHGHAWYKKQATATAEISANTGANTVATVYVSDGGQWTKD----- 66
Db  48  LYDEGGKRVQLKGVSSHQMGFGD-----TYKKDSKMWLRDWMGINV 89
QY  67  -----IHTVRLTISLAEDNHLVAPEVEDHATGYDSTIASLNRAVDWI 108
Db  90  FRVAVYTAADGVYISNPLANKYKAEVAQAAGSLGYIIIDMHLTSDNDPRITYAQAKTPFA 149
QY  109  EKRSLIKGKEDTVITINIANEFQ--SWEGD--AMADGYKQALPRLNAGLNTHTLVADAAG 164
Db  150  EK-AELGYSSPNVYIEINENPGVGTWNGQIRPYA---LEVDITRSKDPDMLITV--G 202
QY  165  WQGFQSHDYCREVFENADPQNTRETSIMHEVYAGAGNSQVATNTDRLVNDLALVIGEF 224
Db  203  TGTWQDIDH---AADNQLPDPNTITVLAHF--YAGTHGQFLDRIDVAYQSGAALFVSEW 257

```

```

OY 225 GRRHNGD-----VEATIMSYSEORGNTLMSKNGNPEMEYL-----DLSDNAGNLT 275
Db 258 GTSDSAGNGGPPPLPSQIWIWIDFLNKRGVSWVMWSLTDKSEASALAPAKSGGRTQNL 317
OY 276 TANGNTI 262
Db 318 STSGKEV 324

RESULT 8
GUN_CLOAB
ID GUN_CLOAB STANDARD: PRT: 448 AA.
AC P15704:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULASE).
GN EGLA.
OS Clostridium acetobutylicum
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
CC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P262;
RX MEDLINE; 88268074.
RA Zappe H., Jones W.A., Jones D.T., Woods D.R.;
RT "Structure of an endo-beta-1,4-glucanase gene from Clostridium
RT acetobutylicum p262 showing homology with endoglucanase genes from
RT Bacillus spp.";
RL Appl. Environ. Microbiol. 54:1289-1292(1988).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION OF C.ACETOBUTYLICUM IS NOT
CC REQUIRED FOR ACTIVITY.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce,
CC CC or send an email to license@sib-sib.ch).
CC -----
CC CC EMBL; M3311; AAA23230.1; -.
CC CC DR PIR; A27631; A27631.
CC CC DR INTERPRO; IPR001547; -.
CC CC PFAM; PF00150; cellulase; 1.
CC CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC CC KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC CC FT SIGNAL 1 34
CC CC FT CHAIN 35 448
CC CC FT ACT_SITE 175 175 PROTON DONOR (BY SIMILARITY).
CC CC FT ACT_SITE 263 263 NUCLEOPHILE (BY SIMILARITY).
CC CC FT ACT_SITE 263 263
CC CC SQ SEQUENCE 448 AA; 49366 MW; 6EB6AC97C836A598 CRC64.

Query Match 11.08; Score 180; DB 1; Length 448;
Best Local Similarity 22.18; Pred. No. 8.7e-07;
Matches 71; Conservative 59; Mismatches 139; Indels 56; Gaps 14;

OY 8 VSGTLLIYANANPEYMRGI-NHGNAVY-----KQATTAIESIANTGANTYAI 54
Db 49 VVGSQLDCSNSEKPIQILKMSHGLQWYNAFYNDSEKELRDK-----WGVNVIRA 98
OY 55 VL--SDGQWT--KDDITVNLISLADNNHVAPEVHADAGYDSIASLNRAVYWIEM 110
Db 99 AMYNIEGGYIINPSSQGRKIKIKIYVDALIDLNNYVLIIDHILSLDNNPNYIKEDQAKSFQDM 158
OY 111 RSALIGKEDTVIINIANEMFGSWEGDANAVGK---QAIPLRNAGNLHTLMDVDAAGWG 166

```

```

Db 159 AEE-YGKSNVYIECNPNQ---GTMMANDIKPYANTIIIPAIAPIDPNNIIIVGISTWS 214
Oy 167 QEPQSIHDYGREVENADPQR--NTWFSIHTMYEYAGNAGNAGVRYNDRLVNDLALVIGEE 224
Db 215 QDVADLAD-----NPLRYSNIMYTCHE--YAGTHQSLRDJINYMMSGIAPVTEW 264
Oy 225 GHRHTNGD----VDEA-TIMSYSEORGVGWLAMSKGNGPMEYEL---DISNDMAGNML 275
Db 265 GTSDASGNGGPFILDESQKWPDPMASKNISWTNVALCDKSEASALKSGSTTGCTSD 324
Oy 276 TAMGNTIYNGPYGLRSTSL 296
Db 325 TTSGLEFVKKSIGSGNTTSGTS 345

RESULT 9
GUN5_THIEFU STANDARD: PRT: 466 AA.
AC 001786:
DT 01-JUL-1993 (Rel. 86, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4)
DE (CELLULOSE E-5) (CELLULOSE E5).
GN CELE.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
RN
RP SEQUENCE FROM N.A.
RX STRAIN-YX;
RX MEDLINE: 91258320.
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RT Thermomonospora fusca";
RT J. Bacteriol. 173:3397-3407(1991).
RN
RN [2]
RA REVISIONS.
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN
RN [3]
RN SEQUENCE OF 137-142 AND 157-166.
RC STRAIN-YX;
RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
RT "Activity studies of eight purified cellulases: specificity,
RT synergism, and binding domain effects.";
RT Biotechnol. Bioeng. 42:1002-1013(1993).
RL Biotechnol. Bioeng. 42:1002-1013(1993).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- PATHWAY: CELLULOSE DEGRADATION.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L01577; AAC09379.1; -.
CC PIR: C42360; C42360.
CC HSSP: P07986; IEXH.
CC INTERPRO: IPR001547; -.
CC INTERPRO: IPR001919; -.
CC PFAM: PF00553; CBD_2; 1.
CC PFAM: PF00150; cellulase; 1.
CC PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
KM
FT SIGNAL 1 36 ENDOGLUCANASE E-5.
FT CHAIN 37 466

```

```

FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 466 AA; 49800 MW; 1CF0ADFBE2DEF82E CRC64;

Query Match 10 9%; Score 177.5; DB 1; Length 466;
Best Local Similarity 24.8%; Pred. No. 1.4e-06;
Matches 78; Conservative 49; Mismatches 121; Indels 67; Gaps 18;

Oy 8 VSGTLYDANGNPFVARGIN-HGHANYKDOAT-TALEGIA-NTGANTVRI---VLSGQ 61
Oy 173 VCGTQLDEHNGNVPVLRKSTHGIOWFCHLFDSDALAYPMKADIILSMYIDEGYE 232
Db 62 WT-----KDIATVRYNLSLAEDNHLVAVPEVHATGTYDLSA:NRADVIRMSRLIK 117
Db 233 TNPGEFTDRMH---QLIDNATARGLYIVDMHTLTGDPDHYNDRAKXTFAELAQHASK 289
Oy 118 EDVYIINIANENFGSWEGDAD--GY-KOALPRLNAGINHTLVDAAGMGQFPQSIHD 174
Db 290 TN-VLYETANE-----PQGVMSATKSYAEVYIPVIRQDPDSVIYVTRGWSLGVY-EG 343
Oy 175 YGREVENADP--QNTWFSIHTMYEYAGNAGVRYNDRLVNDLAL---VIGFGRHT 229
Db 344 SGPAELIANPYNASNIMYAEHFY-----AASHRDYLNALREASLFPVYVEFETET 397
Oy 230 NGD-----VDEATISYSEORGVGWLAMSK-----GNGPMEYIDLS 267
Db 398 TGDANDPQMDR--YIDLMAERKIGTKWNYSDDFRSGAVPQRCASGP----- 447
Oy 268 NDWAGNNLTAWGNTI 282
Db 448 --WSGSLKASGQWV 460

RESULT 10
GUNW_ERWCA STANDARD: PRT: 504 AA.
AC 059395:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ENDOGLUCANASE V1 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V1)
DE (CELLULOSE V1).
GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN-SCC3193;
RX MEDLINE: 95231512.
RA Mae A., Helkinheimo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
RT phytopathogenicity";
RT Mol. Genet. 247:17-26(1995).
RL
RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X79241; CA555823.1; -.
CC HSSP: O06851; INBC.
CC INTERPRO: IPR001547; -.

```

```

DR INTERPRO: IPR001956; -.
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 504 ENDOGLUCANASE V1.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 504 AA: 54963 MW: 927674781555FA CRC64;

Query Match 10.7%; Score 174.5; DB 1; Length 504;
Best Local Similarity 23.7%; Pred. No. 2.7e-06;
Matches 73; Conservative 43; Mismatches 123; Indels 69; Gaps 14;

QY 13 LYDANGNPPVAMGT-NHGMATKQDATTABEGTANTGANTVRIYLSGGQTKDD--IHT 69
DB 48 LYDEGKRYQLRGISNGIQWGD-----YVNRDSKMTLRDDMGIV 89
QY 70 VNLISAEENHLVAVPEY-----HDATGYDSIASLNRAYDYM 107
DB 90 FVAVATTAE-NGYIANPSLANKYKEAVAAAGLGYITIIDMTLSNDNTYKAQKTF 148
QY 108 IEMRSALIGKEDTVIINIANEMFGS--WEGD--AMADGYKQAIPLRNAGLNTLWADA 163
DB 149 AAM-AGLYNSPNVIEIANEPNGSVTNGQIRPYA---LEWTDTRSDPDLNLIYVSG 204
QY 164 GNGQPPQSHDYGREYFNADPQATWESIMTEYAGNAGNSQVNTINDRLNDDLAIVCE 223
DB 205 TWS---ODIHD---AADNLPDPNTLYALHF--YAGTHQGFDRIDYQSGAALFVSE 256
QY 224 FGHRTNGD-----VDEATINSYSEQGYGLWMSKNGKPEMEYL---DISNDAGNN 274
DB 257 WTSIDASGGGFPPLPESQTWIDFLNNRGLISWVMSLSKSETSAAVLVAGASGGWTEON 316
QY 275 LTWAGNTI 282
DB 317 LSTSGKEV 324

RESULT 11
GUN1_BACSU STANDARD; PRT; 499 AA.
AC P07983;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN BGIC OR GLD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-DIG.
RX MEDLINE: 87194581.
RA Robson L.M., Chambliss G.H.;
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DIG.";
RL J. Bacteriol. 169:2017-2025(1987).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16185; AAA22496.1; ALT_INIT.
DR PIR: A26874; A26874.
DR HSSP: 006851; INRC.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29 ENDOGLUCANASE.
FT CHAIN 30 499 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 169 169 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 257 257 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA: 55187 MW: 33904855A63EE1 CRC64;

Query Match 10.6%; Score 173.5; DB 1; Length 499;
Best Local Similarity 20.8%; Pred. No. 3.2e-06;
Matches 69; Conservative 68; Mismatches 122; Indels 73; Gaps 15;

QY 3 NGEFVSGTLYDANGNPPVAMGT-NHGMATKQDATTABEGTANTGANTVRIYLSGGQ 61
DB 39 NQLSISIKGTQVNRDQKAVQLGKISSHGLQWYGD-----FVNRDSIK 80
QY 62 WTKDD--IHT-----VRLISAEENHLVAVPEYHATGYDSIA 98
DB 81 WLRDPMGITVFPAAWTADGTYIDNPVKNKYKEAVEAKELGYITIIDMTLNDGNPQ 140
QY 99 SLNRADVIEKRSALIGKEDTVIINIANEMFGS--WEGD--AMADGYKQAIPLRNAGL 154
DB 141 MEKAKKEFKEM-SSLYGTPVIVIEIANEPNGDVNMKMDIRPYA---EVSIVRKNDP 196
QY 155 NHTLWADAAGWQFPQSHDYGREYFNADPQ--RNTWESIMTEYAGNAGNSQVNTINDR 211
DB 197 DWIITV---GTGWSQDVND-----AADQLKDNVMTALHF--YAGTHQGLDRKANY 245
QY 212 VLNODLAVIGFGRHRTNGD-----VDEA-TIMSYSEQGYGLWMSKNGKPEMEYL-- 264
DB 246 ALSKAPLFTVMTGSDASNGGVFLDOSREWLNTDSKNISWVMSLSKSESSALKP 305
QY 265 --DLSNDVAGNNLTWAGNTIVNGPYGLRSTR 294
DB 306 GASKTGMPDLTDLTASGTFVRENIGTKDSTR 337

RESULT 12
GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCSE) (CELLULOSE).
GN BGIC OR GLD OR BGLS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-PAP15.
RX MEDLINE: 8706793.
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-CK-2;
RX MEDLINE: 95225655.
RA Lindahl V., Aa K., Tromsø A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
  subtilis CK-2."
RL Antonic van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Rose M., Entian K.;
RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 30-45.
RX STRAIN-CK-2;
RC MEDLINE: 95225655.
RA Aa K., Flensburg R., Lindahl V., Tromsø A.;
RT "Characterization of production and enzyme properties of an
  endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
  compost soil."
RL Antonic van Leeuwenhoek 66:319-326(1994).
RN [5]
RP CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
  LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
  HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z29076; CA82317.1; -
DR EMBL: X04689; CA28392.1; -
DR EMBL: X67044; CA47429.1; -
DR EMBL: Z73234; CA97610.1; ALT-INT.
DR EMBL: Z99113; CAB13696.1; ALT-INT.
DR PIR: A26114; A26114.
DR HSP: 006851; INBC.
DR SUBTILIST; BG10A37; BGIC.
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR001956; -
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EAEZ CRC64;

Query Match 10.0%; Score 162.5; DB 1; Length 499;
Best Local Similarity 21.0%; Pred. No. 2.2e-05;
Matches 66; Conservative 75; Mismatches 136; Indels 37; Gaps 15;

```

```

DB 212 ND-----AADOLKDNAMVALHF--YAGTGOFLRKANALSKAPITVTEGTSDA 263
QY 230 NCD-----VDEA-TIMSYSHORGWMLAMSKNGCEWEYL-----DLSNDWAGNNLTJANGN 280
DB 264 SGNGVFLDQSRWELKYDLSKITSIMWNWMLSDKQSSSLAKGASKTGWRSLDLSAGT 323
QY 281 TIVNGPYGLRETSR 294
DB 324 FVENILGTROSTK 337

RESULT 13
GUNA_STRLI STANDARD; PRT; 459 AA.
ID GUNA_STRLI
AC P27035;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ENDOGLUCANASE CELA PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN CELA.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
CC [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 28-43.
RP STRAIN-66;
RC MEDLINE: 92246492.
RA Theberge M., Lacaze P., Shareck F., Morosoli R., Kluepfel D.;
RT "Purification and characterization of an endoglucanase from
  Streptomyces lividans 66 and DNA sequence of the gene."
RL Appl. Environ. Microbiol. 58:815-820(1992).
RN [2]
RP CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
  LINKAGES IN CELLULOSE.
CC -1- P1M: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
  SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
  (CBD).
CC -----
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
  HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M82807; -; NOT_ANNOTATED_CDS.
DR HSP: P07986; IEXH.
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR001919; -
DR PFAM: PF00553; CBD_2; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 27
FT CHAIN 28 459
FT DOMAIN 136 147
FT DOMAIN 148 357
FT ACT_SITE 148 357 CATALYTIC.
FT ACT_SITE 286 286 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 31 131
SQ SEQUENCE 459 AA; 48663 MW; CFA47EC34E2F97A7 CRC64;

Query Match 9.98%; Score 161; DB 1; Length 459;
Best Local Similarity 25.0%; Pred. No. 2.6e-05;
Matches 81; Conservative 42; Mismatches 145; Indels 56; Gaps 16;

```


DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KM Cellulose degradation: Hydrolase; Glycosidase; signal.
 FT SIGNAL 1 29
 FT CHAIN 30 499
 FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 SQ SEQUENCE 499 AA; 55169 MW; 2E821E3D8BAC404 CRC64;

Query Match 9.38; Score 151.5; DB 1; Length 499;

Best Local Similarity 20.78; Pred. No. 0.00016;
 Matches 65; Conservative 75; Mismatches 137; Indels 37; Gaps 15;

```

QY 3 NSGFVSGTLYDANGNFVNRGI-NHGHAWYKDOAT-TAIEGIANTGANTV---RIVLS 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 39 NGQLSIKGTQLYNRDGRAVOQLKGISSHQLQWYGEVNNDSIKWLRDQGLVFRAMAYTA 98
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 DGGQWTKDDI-HTVNNLSLAEDNHLVAVPEVHDATGDTASTASLNRAVDYTIEMRSALIG 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 99 DGGIIDNFSVKKMKKEAVEAKELGIYIIDWHILNDGNPNQNKKEAKEFEKEM-SSLYG 157
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 117 KEDYIITIANEMFG--SMEDG--AMADGYKQALPRLNAGLNHTIATVDAAGWGQFPOSI 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 158 NTPNVIYEIANEFNGDVNMKDKIPYAE--EVLIVIRKNDPNITIV---GTGWSQDV 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 173 HDYGREVFNADPQ---RNTMFSIHMYEYAGGNASQVRTNIDRYLNQDLALVIGFGRHT 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 212 ND-----AADDQLKDAVNDALHF--YAGTRHGFRLDKANAYALSKGAPIFVTEWGTSDA 263
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 230 NCD----VDEA-TIMSYEQRGVGLAWSKRGNGPEWEYL---DLSNDMAGNNLTAMGN 280
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 264 SGNQGVFFLDQSRKWLKYLDSKTIISVWNMNLSDKQSSSALKFGASKTGQWRLSDLSASGT 323
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 281 TIVNGPYGLRSTSR 294
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 324 FVRENILGTRKDKSTK 337
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: December 19, 2000, 16:44:17
 Job time: 893 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:17:09 ; Search time 363.46 Seconds
(without alignments)
77.071 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGFYVSGTLYDANGNP.....TIVNGPYGLRHSRLSTVPT 300

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues
Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_14:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1172	71.9	516	2	066185 bacillus ci
2	797	48.9	395	2	069347 vibrio sp.
3	710	43.6	384	2	086599 streptomyce
4	634	38.9	279	2	092F13 thermomonos
5	614	37.7	930	2	09RFX5 corynebacteri
6	536.5	32.9	327	2	09R115 streptomyce
7	226	13.9	1097	2	09ZAI7 anaerobact
8	197	12.1	557	2	094622 clostridium
9	191	11.7	389	2	059232 bacillus sp
10	176.5	10.8	501	2	083012 bacillus su
11	174.5	10.7	499	2	045532 bacillus su
12	170.5	10.5	930	2	059280 clostridium
13	166.5	10.2	387	2	031029 erwinia car
14	162.5	10.0	486	2	043430 bacillus sp
15	161.5	9.9	499	2	052731 bacillus sp
16	160.5	9.8	481	2	066064 actinomyces
17	159.5	9.8	635	2	066065 fibrobacter
18	157	9.6	783	2	045554 bacillus sp
19	154.5	9.5	749	2	059154 anaerocellu

20	152	9.3	821	2	059241 bacillus sp
21	140	8.6	278	2	09RKS6 streptomyce
22	136.5	8.4	570	2	059665 pseudomonas
23	129	7.9	478	5	016028 globodera r
24	126	7.7	319	5	018454 heterodera
25	126	7.7	438	3	09URL8 corynebacteri
26	126	7.7	476	5	018453 heterodera
27	125	7.7	395	5	09T5M4 globodera t
28	123	7.5	319	5	061595 heterodera
29	123	7.5	319	5	077449 heterodera
30	121.5	7.5	1000	2	024820 thermophil
31	117	7.2	391	5	077094 globodera r
32	117	7.2	392	5	044078 globodera r
33	116.5	7.1	332	2	060054 unidentified
34	116	7.1	426	2	09REW0 erwinia chr
35	115	7.1	363	2	007652 celivibrio
36	114	7.0	419	3	012539 agardicus b1
37	114	7.0	470	5	0906M5 globodera t
38	113	6.9	910	3	087211 orphomyces
39	112.5	6.9	621	2	007653 celivibrio
40	111.5	6.8	506	5	090A57 meloidogyne
41	110	6.7	435	3	012626 picchia angu
42	109.5	6.7	754	2	085318 salmonella
43	107	6.6	430	3	012712 trichoderma
44	106	6.5	325	3	09Y8B6 emericella
45	105	6.4	456	3	09Y7Y7 schizosacch

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	516 AA.
066185	AC	066185		
ID	066185			
DT	01-AUG-1998 (TIREMBLrel. 07, Created)			
DT	01-AUG-1998 (TIREMBLrel. 07, Last sequence update)			
DT	01-MAY-2000 (TIREMBLrel. 13, Last annotation update)			
DE	MANNAVASE.			
OS	Bacillus circulans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98233274.			
RA	Yoshida S., Sako Y., Uchida A.;			
RT	*Cloning, sequence analysis, and expression in Escherichia coli of a			
RT	gene coding for an enzyme from Bacillus circulans K-1 that degrades			
RT	guar gum."			
RL	Biosci. Biotechnol. Biochem. 62:514-520(1998).			
DR	EMBL; AB007123; BAA25878.1; -			
DR	INTERPRO; IPR001547; -			
DR	PFAM; PF00150; cellulase; 1.			
SO	SEQUENCE 516 AA; 55245 MW; D8E5A48842AE8062 CRC64;			

Query Match 71.9%; Score 1172; DB 2; Length 516;
Best Local Similarity 69.2%; Pred. No. 2e-84;
Matches 207; Conservative 50; Mismatches 42; Indels 0; Gaps 0;

QY	1	NANSGFYVSGTLYDANGNPVMEGHNHAWYDQATTAEGTANTGANTVRIVLSDCG	60
DB	33	HAASGFYVSGTLYDANGNPVMEGHNHAWYDQATTAEGTANTGANTVRIVLSDCG	92
QY	61	QWTEDDIHTVNNLSLADNHLVAVPEVHDATGYDSTASINRAVDYTEMSSIGREDT	120
DB	93	KWTDSDVTVNNLTLDEQNKLLAVLEVDHATGSDSLDNVAVYIGIKSALIGREDR	152
QY	121	VVINIANEWFSGWEDAMADGYKQAIPIRLNAGINHTLWDAAGWGFOPSHDYGREVF	180
DB	153	VVINIANEWFSGWEDAMADGYKQAIPIRLNAGINHTLWDAAGWGFOPSHDYGREVF	212
QY	181	NADPQRTMFSIMHYETAGAGNAGVGTINIDRYLNMQDLALVIGEGHRTNGDVDEATIMS	240

```

DB 213 NADLAKTIVEIHHYETAGNASTYKSNIDVLAKNLALIIIGEGCGITNDVDVEATIMS 272
QY 241 YSEORGVGLWAMSKNGPEWEYLDLSNDMAGNNLTANGNTIVNGPYGLRETSRLSTVF 299
DB 273 YSGERGVLWAMSKNGSDLAIDLMDTMDMAGNSLTSPGNTVNGSNGIKATSVLSGIF 331

```

```

RESULT 2
069347 PRELIMINARY; PRT; 395 AA.
AC 069347;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE BETA-1,4-MANNANASE.
GN MANA.
OS Vibrio sp.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MA-138;
RA Tamaru Y., Ataki T., Morishita T., Kimura T., Sakka K., Ohmura K.;
RL J. Ferment. Bioceng. 83:201-205(1997).
DR EMBL; D86329; BAA25188.1; -
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR002883; -
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 395 AA; 43098 MW; A3E30B3A978C5DA1 CRC64;

```

Query Match 48.98; Score 797; DB 2; Length 395;
 Best Local Similarity 50.88; Pred. No. 4.4e-55;
 Matches 155; Conservative 55; Mismatches 87; Indels 8; Gaps 5;

```

QY 2 ANSGFYVSGTLLDANGNPFVNRGINHGHWYKQDQATTAIGTANTGANTVRYVLSGQ 61
DB 19 AHAFYVSNGLVETANGSAFKIRGINHATHTDGLSVLALSGIATGANTVRYVLSNGYR 78
QY 62 WTKDHIHVRNLISLAEDNHLVAVPEVDATGY--DSIASINRAVDYIEMRSALIGKE 118
DB 79 WTKDSDVNTINILAKANNILALIEVDHTGYGESSAALSADADWIELKELIGQE 138
QY 119 DTVIINTANEFQ-SWEDANADGYKQAIPLRNAGLHNTLVADAGGQ-FPOSINDY 176
DB 139 DVIINLGNPEFGNNNDVAVVNDHVSRIORLSGINHTIWDAPNMGODWKGFMNNA 198
QY 177 REVNADPQRTMFSIHYE-YAGGNASQVRNIDRVLNODLALVIGFGRHRTGVDYE 235
DB 199 QFVNSDPKLTIPSVHMYEYTSYNS--VNDYISFTNNGLVLYIGFASHTKCADYE 256
QY 236 ATIMSYSEORGVGLWAMSKNGPEWEYLDLSNDMAGNNLTANGNTIVNGPYGLRETSRL 295
DB 257 GSIMERSSETLSGLYIGWMSGNDTTSDDLTYNNMNDNNSYSTGWGLLNGNGISTSTL 316
QY 296 STVFT 300
DB 317 ATVFT 321

```

```

RESULT 3
086599 PRELIMINARY; PRT; 384 AA.
AC 086599;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE SECRETED BETA-MANNOSIDASE.
GN MANA2.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

```

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Randleam M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach K., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031514; CA20610.1; -
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR002883; -
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 384 AA; 39847 MW; B8AD1583F3BD11 CRC64;

```

Query Match 43.68; Score 710; DB 2; Length 384;
 Best Local Similarity 46.74; Pred. No. 3e-48;
 Matches 141; Conservative 53; Mismatches 102; Indels 6; Gaps 4;

```

QY 2 ANSGFYVSGTLLDANGNPFVNRGINHGHWYKQDQATTAIGTANTGANTVRYVLSGQ 61
DB 37 AAGGIHVSNGVLYGNSGVYFMRGVNHYWYPPDR-TGSIADIAKAGANTVRYVLSGGR 95
QY 62 WTKDHIHVRNLISLAEDNHLVAVPEVDATGY--DSIASINRAVDYIEMRSALIGKE 118
DB 96 WTKSASEVSNALIOCCANKNVICVLEVDHTGYGEGDAATSLDQADYVSVSALGQE 155
QY 119 DTVIINTANEFQSWEDANADGYKQAIPLRNAGLHNTLVADAGGQ-FPOSINDY 177
DB 156 DVIYVNVGNPEFGNTANTATDKASIGLKRGLDHALVADAPNMGODWKGFMNNA 215
QY 178 EVENADPQRTMFSIHMYETAGNASQVRNIDRVLNODLALVIGFGRHRTGVDY 237
DB 216 SVFASDDPRTMFSVHYG-VYDFAAEVRDYLNALFVSGLPYVGERGDHSDENPEDA 274
QY 238 IMASSEORGVGLWAMSKNGPEWEYLDLSNDMAGNNLTANGNTIVNGPYGLRETSRLST 297
DB 275 IMATQSLGVGLYLSGWSGNGGVYLDVNVGFPDNLSTGNRIFYGSNGIAATSTRTAT 334
QY 298 VF 299
DB 335 VY 336

```

```

RESULT 4
092F13 PRELIMINARY; PRT; 279 AA.
AC 092F13;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE BETA-MANNANASE (EC 3.2.1.78) (FRAGMENT).
GN MAN.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KW3;

```

RA Hluge M., Gloor S.M., Winterhalter K., Zimmermann W., Piontek K.;
 RT "Crystallization and preliminary crystallographic analysis of two
 RT beta-mannanase isoforms from *Thermomonospora fusca* KM3.";
 RL Acta Crystallogr. D 52:1224-1225(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-KW3;
 RC MEDLINE: 99036862.
 RA Hluge M., Gloor S.M., Rypniewski W., Sauer O., Reichtman T.D.,
 Zimmermann W., Winterhalter K., Piontek K.;
 RT "High-resolution native and complex structures of thermostable beta-
 RT mannanase from *Thermomonospora fusca* - substrate specificity in
 RT glycosyl hydrolase family 5.";
 RL Structure 6:1433-1444(1998).
 CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC
 CC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
 CC GALACTOGLUCOMANNANS.
 CC EMBL: AJ006227; CA006924.1; -.
 DR INTERPRO: IPR001547; -.
 DR PFAM: PF00150; cellulase: 1.
 DR PROSITE: PS00659; GLYCOSTYL_HYDROL_F5; UNKNOWN_1.
 KW Hydrolyase; Glycosidase.
 FT NON_TER 1 1
 FT 279 279
 SQ SEQUENCE 279 AA; 30657 MW; 98A2850FAE013DB0 CRC64;

Query Match 38.9%; Score 634; DB 2; Length 279;
 Best Local Similarity 44.4%; Pred. No. 1.8e-42;
 Matches 128; Conservative 56; Mismatches 82; Indels 22; Gaps 6;

QY 5 GFVSGTLYDANGNPPVAMGNGHAYKQDTAEGTANTGATVRYVSDGGQWTK 64
 DB 1 GLTVKNGRYLEANGQEEIIRGVSHPHNMY-POHTQAPADIKSGANTVRVLSNGVWMSK 59
 QY 65 DDHIVRNLIISLAEDNLVAVPEVHDATGY---DSIASLRAYDYIEMKRSALIGKEDTV 121
 DB 60 NGPSDVANVSLCKONRLICMLTEVHDTGTGEGSGASTLDQADVTELSVYQGEEDYV 119
 QY 122 IINIANEWF-----GSWEGDAMADYKQALPRLNRAGLHNTLVADAAGNQ-PPQS 171
 DB 120 LINGNEPEYNDSATVAGAMDTSA-----AIQRRAAGFEHTLVADPNNGQDWTNT 172
 QY 172 IHDYGEVFADPQRTMFSIHYTEVAGASOVKRTIDVNLQDLALVYGECHRTNG 231
 DB 173 MRNNADQVVASDPTGNVTFIHMVG-VYSGASTTSTYIEHFVNAGPLLIIGEGHDSG 231
 QY 232 DVDDEATIMSYEORGWGLAMSWKNGPEWEYIDLSDMAGNLTLAG 279
 DB 232 NPDEDITMAERLKLGIYGSMSNGGVEYIDMTYNTFDGDLSPWG 279

RESULT 5
 ID 09RFX5 PRELIMINARY; PRT; 930 AA.
 AC 09RFX5;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
 DE MULTIDOMAIN BETA-1,4-MANNANASE PRECURSOR.
 GN MANA.
 OS *Caldibacillus cellulosovorans*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; *Caldibacillus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sunna A., Gibbs M.D., Bergquist P.L.;
 RT "A novel multidomain beta-1,4-mannanase gene from *Caldibacillus*
 RT *cellulosovorans* and action of the recombinant enzyme on krat pulp.";
 RL EMBL: AF163837; AAF22274.1; -.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR001956; -.

DR INTERPRO: IPR002965; -.
 DR PFAM: PF00150; cellulase: 1.
 DR PFAM: PF00942; CBD_3; 2.
 DR PRINTS: PR01217; PRICEXTENS.
 DR PROSITE: PS00659; GLYCOSTYL_HYDROL_F5; UNKNOWN_1.
 KW Signal.
 FT SIGNAL 1 33
 FT CHAIN 34 930
 SQ SEQUENCE 930 AA; 101516 MW; 0086538D54D1A2CC CRC64;

Query Match 37.7%; Score 614; DB 2; Length 930;
 Best Local Similarity 46.9%; Pred. No. 3.9e-40;
 Matches 122; Conservative 52; Mismatches 80; Indels 6; Gaps 5;

QY 25 GINHGAWRYKQDTAEGTANTGATVRYVSDGGQWTKDHTVNTLSLAED-NHLY 83
 DB 486 GINHPHAWRYDRLLSSLOGIRSGANAVRYVLSNGCRWTKRIPASEVADISQARTIGYRA 545
 QY 84 AYPEVHDATGY--DSTA-SINRAVDYIEMKRSALIGKEDTVIINIANEWFGSWEGDAMAD 140
 DB 546 VVLEVHDYTGEGDAAACSTVTVNMYIELKNVLAGQENFVIVNIGNEPYGNNTYQWYV 605
 QY 141 GYKQALPRLNAGLNTIWDADAGNGO-PPQSITHDYGREVENAPQRTMESTIHMYEYAG 199
 DB 606 DTRNAVALNAGINNTINWDAPNMGDWSETEDNAPTFINADPQRTLVFSIHMVG-VY 664
 QY 200 GNASQVNTIDRYLNDLALVYGEFGRHNGDVDEATIMSYEORGWGLAMSWKNGP 259
 DB 665 DTAELVQSYIESYVNGGLPLVYGEFGRHNSDGPNDQALYQAKQINIGLFGMSNGNG 724
 QY 260 EWEYLDLSNDWAGNNTFANG 279
 DB 725 GVEYLDMTYNFNANSPTLAWG 744

RESULT 6
 ID 09RUT5 PRELIMINARY; PRT; 327 AA.
 AC 09RUT5;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
 DE PUTATIVE SECRETED BETA-MANNOSIDASE (FRAGMENT).
 GN MANA.
 OS *Streptomyces coelicolor*.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Rajkhi J., Barrell B.G., Rajandream M.A.;
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleser H.M., Denaparte D., Elchner A., Cullum J.,
 RA Kinash H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb *Streptomyces coelicolor* A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AF121746; CB857406.1; -.
 DR INTERPRO: IPR001547; -.
 DR PFAM: PF00150; cellulase: 1.
 FT NON_TER 327 327
 SQ SEQUENCE 327 AA; 34772 MW; A21B0068C2C0975D CRC64;

Query Match	32.94;	Score 536.5;	DB 2;	Length 327;
Best Local Similarity	41.98;	Pred. No. 1.1e-34;		
Matches 114;	Conservative 52;	Mismatches 97;	Indels 9;	Gaps 7;
Qy 4	SGFVSSGTTLLDANGNPFVAMGKINRHHAWYKDOATTAEGTANTGANTVRIVLSDGGQMT 63			
Db 57	AGLHIGGRLLLEGNGNDPFVAMGVNHAHWYGE -TGSIAVKKALGANSVRVYVLSDGHRMS 115			
Qy 64	KDDIHVKNLLSLAEDNHLVAPEVHDATGY -DSIA-SLNRAVDYITEMRSALIGKEDT 120			
Db 116	ENGPAVVAIVIEQCKANRLICVLEHVDHTGYAEDPAAGTLDHADYWGILGDVYLAGGDXY 175			
Qy 121	VIVIANEMFESSWGADAMADGYKQALPRLRNAGLNTLMDVPAAGCGCPQSI -HDYGRVY 179			
Db 176	VIVINGNPGMNTDPAGTEPTVAAYKKLRAGLOHTLMDVAPMVGQDMGMRANASV 235			
Qy 180	FNAPQQRMTFSIHXYEYAGNVAISOVRNTIRVNLQDIALYIGF -GHRHTNGDYDEATI 238			
Db 236	YDADTGTLIRLSIHMTS -VFDTAQETITLYLNFVAELPILIGERGGRADQYGPDEPDTM 294			
Qy 239	MSYSEQRGVGNLWMSKNGPMEYILDSNDW 270			
Db 295	MATAQLRLGYLWMSGNTD -PYLDALDP 324			
RESULT 7				
Q9ZAL7	PRELIMINARY;	PRT;	1097 AA.	
AC 09ZAL7				
DT 01-MAY-1999	(TREMBLrel. 10, Created)			
DT 01-MAY-1999	(TREMBLrel. 10, Last sequence update)			
DT 01-MAY-2000	(TREMBLrel. 13, Last annotation update)			
DE	S-LAYER ASSOCIATED MULTIDOMAIN ENDOGLUCINASE.			
GN	CELA.			
OS	anaerobic thermophile KM-THCJ.			
NC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-KM-THCJ;			
RX	MEDLINE: 99175466.			
RT	Cann I.K., Kocherginskaya S., King M.R., White B.A., Mackie R.I.;			
RL	"Molecular cloning, sequencing, and expression of a novel multidomain			
RL	mannanase gene from Thermanaerobacterium polysaccharolyticum.";			
DR	EMBL: U82255; AAD09354.1; -			
DR	INTERPRO: IPR001119; -			
DR	INTERPRO: IPR001547; -			
DR	INTERPRO: IPR002088; -			
DR	PFAM: PF00395; cellulase; 1.			
DR	PFAM: PF02018; CBD; 6; 2.			
DR	PROSITE: PS00904; PPRA; UNKNOWN_1.			
DR	PROSITE: PS01072; SLH_DOMAIN; 2.			
DR	SEQUENCE 1097 AA; 119762 MW; C41CEB92C0494B9 CRC64;			

```

Query Match 13.9%; Score 226; DB 2; Length 1097;
Best Local Similarity 24.8%; Pred. No. 1.7e-09;
Matches 83; Conservative 60; Mismatches 134; Indels 56; Gaps 15

Qy 6 FVYSGTLLDANGNPVVMGIN-HGH-AWTKDQATAIEGIANTGA-NTVRYVLSDG--- 59
   | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 41 FHVYSGKRIYDPDQNDPEVINGVNTIGQYRSMEKRSVLDVHLDMVMEKFNFTVLCFIGNN 100
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 60 ---GQTKDQDIHRYRNLISLAEDNHLVAVPEVDADGYSDIA-----SINRAVDW 107
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 101 WEGTGANDIDAI---IAFTAKKVVYEIDLDYTGIPPLSNPPAPGQPSDQAI-AW 156
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 108 IEKRSALIGKEEDVIINIAEMFGSWE--GDAAVDGKQAIPLRLRNAGLHHTLAVDAAG 165
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 157 EKRLAAKYKNDPFRVYNTNMPESSSTAPLDQPKRVANEEIIKRLSGADNIIYVD--GW 214
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 166 GQFPGSIHD-----YGREVNADPQRTMTSIEHTETAYAGNAAQVQRINIDR 211

```

Db	215	SYANEGLQNPPTVDEKRSAVLITGGDLNADSKNITFEFHHYN-EGDLOKKVEVDYDR	273
Qy	212	VLNODLAVLIGEGFHRHTNGVDDEATMTMSE--QRCVGLAWSW-----KGN	257
Db	274	ANAGLVYFMEYKRDYS--DAAEVKGSGIQAAVMNNGAGRIYMNMGYDLVLTISGTR	331
Qy	258	GREPEYLDLSNDMAGNNILTAGNTIVAGPYGLRETT	292
Db	332	GSQWE-INKIDGSKPTNLSWVGDKIIMDNGHITPT	365

```

RESULT      8
P946622
ID          P94622          PRELIMINARY;          PRT;          557 AA.
AC          P94622;
DT          01-MAY-1997 (TREMBLrel. 03, Created)
DT          01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT          01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE          ENDO-1,4-BETA GLUCANASE ENGf (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
DE          (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
GN          ENGf.
OS          Clostridium cellulovorans.
OC          Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC          Clostridium.
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE: 97136705.
RA          Shewetta S.A., Ichii-Ishi A., Park J.S., Liu C., Malburg L.M.,
RA          Doi R.H.;
RT          "Characterization of engf, a gene for a non-cellulosomal Clostridium
RT          cellulovorans endoglucanase."
RL          Gene 182:163-167(1996).
CC          -1- CATALYTIC ACTIVITY = ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC          LINKAGES IN CELLULOSE.
DR          EMBL: U37056; AAA40891.1; -.
DR          HSSP: O85465; 2A3H.
DR          INTERPRO: IPR001547; -.
DR          PfAM: PF00150; cellulase; 1.
DR          PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW          Hydrolase; Glycosidase.
SQ          SEQUENCE 557 AA; 60131 MW; D186EC88EB504EED CMC64;

```

```

Query Match      12.1% Score 197; DB 2; Length 557;
Best Local Similarity 26.7%; Pred.No.1.2e-07;
Matches 88; Conservative 33; Mismatches 132; Indels 76; Gaps 17.

QY      12 TYDANGNPFVARGIN-HEHAWYKQATTAIGIANTGA-----NTVAIVL--SD 58
        || :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      62 TCDKDGNFIQLRGKSTHLOMF-----PGVVNNNAFALSNOMSNVIRIAMVAE 113

QY      59 GGQWTKDDI-HFYRNIIISIAEDNHLYAVEPHDAICYSIASLNRAVDYIEKRSALLGK 117
        || :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      114 GGATNPVKQCFYINGINALANDMYAIDWHMNMGPDENASYSQAQSEFNDSITLYPN 173

QY      118 EDYVIINIINEMGWSG---DA-WADQKOAI-P-BLRNAGNLHTLVADAAGGOEP 169
        || :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      174 NKNIIIELCPEPENGNGGVTTNDTGAOYKSYATPIYOLLRKNGENLLIIVGNPFNSQRP 233

QY      170 QSIHGYGREVFNAPOPMTFESIIMETVAGN-AQSVRTNDR-----VLNODIALVT 221
        || :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      234 DLAAD-----NFINSNTMISYHF--YSTGNTISTVDNRBMAASNVRXLAHGAAYFA 285

QY      222 GFEGHRHTNGDVDEATIMTSYGORGCVGLAMSKGSGPEKEIYDLND----- 269
        || :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      286 TEMGTSLATG----TTGPYL-AKADAWDF-LNGNNISWCNESISKDEKAALNSLTGS 338

QY      270 -----WAGNNLTAMGNTI---VNGPY 287
        || :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      339 LDPSGDKLMADNELLTSGQTVARIRIGAY 367

```


DB 39 NGOLSTKQVLYNRDGRKAVOLKGISSHGLQWYGD-----FVNKSLK 80
 QY 62 WTKD-----IHT-----VRNLISLAEDNHLVAVPEVDATGYSIA 98
 DB 81 WLRDWMGIVTFRRAKTYTADGGYIDNP SVKRNKKEAVEAKELGIVTIDMHIIDNDGNPQ 140
 QY 99 SLNRVADVWEIRSAIIGKEDVTIINIANEMFG--SWECD--ANADGYKQAIPLRNAGL 154
 DB 141 HEKAKDFPEK--SLIKGTPTVITIEINERPGDYNMRDIPYAE--EYISVIRKNDP 196
 QY 155 NHTLWADAGWQFPOSIDHYREVFNADPO--RNFESTIMTYAGNAGNAGVPTINDR 211
 DB 197 DNIITV---GTGTWSQDVND-----AADDLKANDVYALHF--YAGTHGQSLDKRANY 245
 QY 212 VANOALVYIGFGRHNTGD-----VDEA--TIMSYSEORGWGLAWSKNGPEWEYL-- 264
 DB 246 ALSKAPITVTEGTSIDAGNGGVFLDOSREWLNTIDSKNISVWNNLSDKQESSALKP 305
 QY 265 --DLSDMAGNNLTWANGTIVNGPYGLRSTR 294
 DB 306 GASKTGWPLTDLTASTGTFRNLTGNKDKSTK 337

RESULT 12

059290 PRELIMINARY; PRT; 930 AA.
 AC 059290;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE BETA-1,4-ENDOGLUCANASE (EC 3.2.1.4).
 GN CELA.
 OS Clostridium josui.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fujino T., Fujino E., Karita S., Ohmura K.;
 RT "Revised sequence of cels gene encoding endoglucanase (Eg)-1 from
 RT Clostridium josui."
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D85526; BAA12826.1; -.
 DR HSSP; O85465; 2A3H.
 DR INTERPRO; IPR001119; -.
 DR INTERPRO; IPR001547; -.
 DR PFAM; PF00150; cellulase; 1.
 DR PRAM; PF00395; SLH; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 930 AA; 101726 MW; 0EBAE8754D084448 CRC64;

Query Match 10.5%; Score 170.5; DB 2; Length 930;
 Best Local Similarity 26.0%; Pred. No. 3.2e-05;
 Matches 76; Conservative 45; Mismatches 106; Indels 71; Gaps 19;

QY 12 TLVDANGNPFVARGIN-HGHAMT-----KQATTLIEGANTGANTVATLV--SDGGQWTK 64
 DB 61 TLCDKDGNDIOLRGMTSGIOMFPEIINNNAFALS--KDMGSVIVRIAMTVAAEGGSKD 118
 QY 65 DDIHIVRNL--ISLAEDNHLVAVPEVDATGYSIASLNR--AVDYWEIRSAIIGKEDVT 121
 DB 119 PELIKRVIDGIDLAINDMVIYIDWHVLTIPGDPNADYTKGADPFKEI--SQKTPNPHI 177
 QY 122 IINIANEMFGSWEF--DA--WA--DGYKQAIPL--LRNAGLHTLWADAGWQFPOSIDH 173
 DB 178 IYELANESPNDPGYTNDAAAGAKYKSYAEPIIKITLRDSGNKMLITVSGPMSQRP---- 233
 QY 174 DYGREVFNADPQRTMTFSIHMTETIYAGNAGNAGVPTINDRVLNODLALVYIGFGRHNTGDV 233
 DB 234 DLAAE--NPINDNNATVSPHFYSGT-----HKTSTDST 264

QY 234 DEATIMS---YSEORGWGLAWSW-----KGNP-----EM-EYLDLSN--DNAGNNLT 276
 DB 265 DRGNINSMARVIALEHGVAVFCSEWGTSEASGNNGFYLKEADWLEFLIANNISWLNWSLT 324

RESULT 13

031029 PRELIMINARY; PRT; 387 AA.
 ID 031029;
 AC 031029;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
 GN CELA.
 OS Erwinia carotovora subsp. carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwinia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LY34;
 RA Park Y.W., Lim S.T., Yun H.D.;
 RL Mol. Cells 0:0-0(1997).
 DR EMBL; AF025768; AAC02964.1; -.
 DR HSSP; O85465; 2A3H.
 DR INTERPRO; IPR001547; -.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 31
 FT CHAIN 32 387
 FT FT 387
 SQ SEQUENCE 387 AA; 42003 MW; 6302B3DF8AC2738B CRC64;

Query Match 10.2%; Score 166.5; DB 2; Length 387;
 Best Local Similarity 21.5%; Pred. No. 1.8e-05;
 Matches 68; Conservative 50; Mismatches 113; Indels 85; Gaps 13;

QY 13 LYDANGNPFVARGI-NHGHAMTKDQATTAIEGANTGANTVATLVSDGGQWTKD----- 66
 DB 48 LYDEGRKRVOLRGVSSHGLQWFGD-----YKKRDSKMWLPDQWGINV 89
 QY 67 -----IHTVRNLISLAEDNHLVAVPEVDATGYSIASLNRVADYWI 108
 DB 90 SRVANTYADGYISKPSLIANKVKEVAAQSLGYITIIDMHIISDNENPTIKKEAKTFEFA 149
 QY 109 EKRSAIIGKEDVTIINIANEMFG--SWECD--ANADGYKQAIPLRNAGLHTLWADAG 164
 DB 150 EM-AGLYGNSPVIYIEINERPGDYNMRDIPYAE--LEVETIRSRKDPNLTIV--G 202
 QY 165 WQFPOSIDHYREVFNADPQRTMTFSIHMTETIYAGNAGNAGVPTINDRVLNODLALVIGEF 224
 DB 203 TGTWSQDIDHD--AADNOLPDPNTLYALHF--YAGTHGQFLIRIDVQAQSGAALIFVSEW 257
 QY 225 GHRHNTGD-----VDEATIMSSEORGWGLAWSKNGPEWEYLDLS----- 267
 DB 258 GTSIDAGNGGVFLDOSREWLNTIDSKNISVWNNLSDKQESSALKP 305
 QY 268 -NDMAGNNLTWANGTIV 282
 DB 309 CGGWTGQNLNAGSKFV 324

RESULT 14

045430 PRELIMINARY; PRT; 486 AA.
 ID 045430;
 AC 045430;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CARBOXYMETHYL-CELLULOSE) (CMCSE) (CELLULOSE).
 GN CELA.
 OS Bacillus sp.

OC	Bacterifirmicutes; Bacillus/Clostridium group;
CC	Bacillus/Staphylococcus group; Bacillus.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-D04;
RX	MEDLINE; 96029707.
RA	Han S.J., Yoo Y.J., Kang H.S.;
RT	"Characterization of a bifunctional cellulase and its structural gene
RT	the cell gene of Bacillus sp. D04 has exo- and endoglucanase
RT	activity.";
RL	J. Biol. Chem. 270:26012-26019,(1995).
CC	-1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC	LINKAGES IN CELLULOSE (BY SIMILARITY).
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC	HYDROLASES).
DR	EMBL; U27084; AAC43478.1; -.
DR	HSSP; 085465; 2A3H.
DR	INTERPRO; IPR001547; -.
DR	INTERPRO; IPR001956; -.
DR	PFAM; PF001150; cellulase; 1.
DR	PFAM; PF00942; CBD_3; 1.
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
RR	PRODOM; PD001947; -; 1.
KM	Signal.
FT	SIGNAL
FT	1
FT	29
FT	POTENTIAL.
FT	ENDOGLUCANASE.
FT	CHAIN
FT	30
FT	486
FT	ENDOGLUCANASE.
FT	DOMAIN
FT	350
FT	486
FT	CELLULOSE-BINDING (BY SIMILARITY).
FT	ACT_SITE
FT	169
FT	257
FT	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE
FT	257
FT	NUCLEOPHILE (BY SIMILARITY).
QO	SEQUENCE
QO	486 AA; 53820 MW; BD3B65BF18D6656B CRC64;

[illegible]

CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Jung K.H., Chun Y.C., Lee J.-C., Kim J.H., Yoon K.-H.;
 RL Biotechnol. Lett. 18:1077-1082(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Yoon K.-H., Jung K.H.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ENDOPHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: AF045482; AAC02536.1; -.
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR001956; -.
 DR PRAM: PF00150; cellulase; 1.
 DR PFAM: PF00942; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM: PD001947; -; 1.
 KW Hydrolase; Glycosylase.
 SQ SEQUENCE 499 AA; 55193 MW; 990A18786CF738 CRC64;

[illegible]

Search completed: December 19, 2000, 16:17:12
Job time: 1013 sec

Wed Dec 20 15:36:58 2000

us-09-339-159-2_copy_31_330_1.rspt

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)